

Education

How the world's
most improved
school systems
keep getting
better. 😊

The complete report can be found at:

<http://mckinseysociety.com/how-the-worlds-most-improved-school-systems-keep-getting-better/>

Executive Summary

How does a school system with poor performance become good? And how does one with good performance become excellent? These were the questions policymakers and education leaders asked us in the wake of our 2007 report *How the World's Best Performing School Systems Come Out on Top*, in which we examined the common attributes of high-performing school systems.

In this new report, *How the World's Most Improved School Systems Keep Getting Better*, we attempt to answer these questions. We analyzed twenty systems from around the world, all with improving but differing levels of performance, examining how each has achieved significant, sustained, and widespread gains in student outcomes, as measured by international and national assessments. Based on over 200 interviews with system stakeholders and analysis of some 600 interventions carried out by these systems – together comprising what we believe is the most comprehensive database of global school system reform ever assembled – this report identifies the reform elements that are replicable for school systems elsewhere as they move from poor to fair to good to great to excellent performance.

The systems we studied were: Armenia, Aspire (a US charter school system), Boston (Massachusetts), Chile, England, Ghana, Hong Kong, Jordan, Latvia, Lithuania, Long Beach (California), Madhya Pradesh (India), Minas Gerais (Brazil), Ontario (Canada), Poland, Saxony (Germany), Singapore, Slovenia, South Korea, and Western Cape (South Africa).

The report's findings include the following eight highlights:

¹ A system can make significant gains from wherever it starts – and these gains can be achieved in six years or less. Student outcomes in a large number of systems have either stagnated or regressed over the last ten years. However, our sample shows that substantial improvement can be achieved relatively quickly. For instance, Latvian students in 2006 demonstrated performance that was half a school-year advanced to that of students in 2000. In Long Beach, six years of interventions increased student performance in grade four and five math by 50 percent and 75 percent respectively. Even systems starting from low levels of performance, such as Madhya Pradesh in India, Minas Gerais in Brazil, and Western Cape in South Africa, have significantly improved their literacy and numeracy levels within just two to four years, while making strides in narrowing the achievement gap between students from different socio-economic backgrounds. Improvement can start from any student outcome level, whatever the geography, culture or income.

2 There is too little focus on 'process' in the debate today. Improving system performance ultimately comes down to improving the learning experience of students in their classrooms. School systems do three types of things to achieve this goal – they change their *structure* by establishing new institutions or school types, altering school years and levels, or decentralizing system responsibilities; they change their *resources* by adding more education staff to schools or by increasing system funding; and, they change their *processes* by modifying curriculum and improving the way that teachers instruct and principals lead. All three of these intervention types – structure, resources, and process – are important along the improvement journey. The public debate, however, often centers on structure and resource due to their stakeholder implications. However, we find that the vast majority of interventions made by the improving systems in our sample are 'process' in nature; and, within this area, improving systems generally spend more of their activity on improving how instruction is delivered than on changing the content of what is delivered.

3 Each particular stage of the school system improvement journey is associated with a unique set of interventions. Our research suggests all improving systems implement similar sets of interventions to move from one particular performance level to the next, irrespective of culture, geography, politics, or history. For example, the interventions undertaken by Madhya Pradesh (India), Minas Gerais (Brazil), and Western Cape (South Africa) on the path from poor to fair performance have striking similarities. There is a consistent cluster of interventions that moves systems from poor performance to fair, a second cluster of interventions

does the same from fair performance to good, a third cluster from good performance to great, and yet another from great performance to excellent. For example, systems moving from fair performance to good focused on establishing the foundations of data gathering, organization, finances, pedagogy, while systems on the path from good performance to great focused on shaping the teaching profession such that its requirements, practices, and career paths are as clearly defined as those in medicine and law. This suggests that systems would do well to learn from those at a similar stage of the journey, rather than from those that are at significantly different levels of performance. It also shows that systems cannot continue to improve by simply doing more of what brought them past success.

4 A system's context might not determine *what* needs to be done, but it does determine *how* it is done. Though each performance stage is associated with a common set of interventions, there is substantial variation in how a system implements these interventions with regard to their sequence, timing, and roll-out – there is little or no evidence of a "one-size-fits-all" approach to reform implementation. Our interviews with system leaders suggests that one of the most important implementation decisions is the emphasis a system places on *mandating* versus *persuading* stakeholders to comply with reforms. For example, while all improving systems make substantial use of data to inform their reform programs, only a subset of our sample systems translate this into quantitative targets at both school and classroom level, and then share this information publicly (U.S., England, Canada, Madhya Pradesh, and Minas Gerais). In contrast, Asian and Eastern European systems

refrain from target-setting and only make system-level data available publicly. Instead, they prefer to share performance data with individual schools, engaging them in a private dialogue about how they can improve. The systems we studied have adopted different combinations of mandating and persuading to implement the same set of interventions. For example, a system will tend towards persuasion when there are stark winners and losers as a result of the change, it can afford a longer implementation time-line, the desired change is not a precursor for other changes, the system and national leadership is at a tenuous moment of credibility and stability, and/or the historical legacy of the nation makes enforcement of top-down decisions difficult.

5 Six interventions occur equally at every performance stage for all systems. Our research suggests that six interventions are common to all performance stages across the entire improvement journey: building the instructional skills of teachers and management skills of principals; assessing students, improving data systems, facilitating improvement through the introduction of policy documents and education laws, revising standards and curriculum, and ensuring an appropriate reward and remuneration structure for teachers and principals. Though these interventions occur at all performance stages, they manifest differently at each stage. Taking the example of teacher training, for instance: while Armenia (on the journey from fair to good) relied on centrally-driven, cascaded teacher training programs, Singapore (on the journey from good to great) allowed teachers flexibility in selecting the topics that were most relevant to their development needs.

6 Systems further along the journey sustain improvement by balancing school autonomy with consistent teaching practice. While our study shows that systems in poor and fair performance achieve improvement through a center that increases and scripts instructional practice for schools and teachers, such an approach does not work for systems in 'good' performance onwards. Rather, these systems achieve improvement by the center increasing the responsibilities and flexibilities of schools and teachers to shape instructional practice – one-third of the systems in the 'good to great' journey and just less than two-thirds of the systems in the 'great to excellent' journey decentralize pedagogical rights to the middle layer (e.g. districts) or schools. However, in parallel, the center mitigates the risk of these freedoms resulting in wide and uncontrolled performance variations across schools by establishing mechanisms that make teachers responsible to each other as professionals for both their own performance and that of their colleagues. For example, these systems establish teacher career paths whereby higher skill teachers increasingly take on responsibility for supporting their juniors to achieve instructional excellence first within the school, then across the system. These systems also establish collaborative practices between teachers within and across

schools that emphasize making practice public – such as weekly lesson-planning for all teachers in the same subject, required lesson observations, and joint-teaching – that serve to perpetuate and further develop the established pedagogy. Although teachers receive 56 percent of all support interventions in our studied systems, they receive only 3 percent of accountability interventions. In other words, collaborative practice becomes the main mechanism both for improving teaching practice and making teachers accountable to each other.

7 Leaders take advantage of changed circumstances to ignite reforms. Across all the systems we studied, one or more of three circumstances produced the conditions that triggered reform: a socio-economic crisis; a high profile, critical report of system performance; or a change in leadership. In fifteen out of the twenty systems studied, two or more of these "ignition" events were present prior to the launch of the reform efforts. By far, the most common event to spark the drive to reform is a change in leadership: every system we studied relied upon the presence and energy of a new leader, either political or strategic, to jumpstart their reforms. New strategic leaders were present in all of our sample systems, and new political leaders present in half.

Critically, being new in and of itself is insufficient for success – these new leaders tend to follow a consistent "playbook" of practices upon entering office to lay the foundations for their improvement journey.

8 Leadership continuity is essential. Leadership is essential not only in sparking reform but in sustaining it. Two things stand out about the leaders of improving systems. Firstly, their longevity: the median tenure of the new strategic leaders is six years and that of the new political leaders is seven years. This is in stark contrast to a norm: for example, the average tenure for superintendents of urban school districts in the U.S. is just three years; the average tenure of education secretaries in England just two years; similarly, that of education ministers in France is two years. Secondly, improving systems actively cultivate the next generation of system leaders, ensuring a smooth transition of leadership and the longer-term continuity in reform goals. This second observation lies at the heart of how a handful of our studied systems (e.g. Armenia, Western Cape, Lithuania) have managed reform continuity despite regular changes of political leadership. The stability of reform direction is critical to achieving the quick gains in student outcomes outlined above.

The fundamental challenge school system leaders face is how to

shepherd their system through a journey to higher student outcomes. This journey is all the more complex because system starting points are different, contextual realities vary, and system leaders face multiple choices and combinations of what to do along the way – a single misplaced step can result in system leaders inadvertently taking a path that cannot get them to their desired destination. While there is no single path to improving school system performance, the experiences of all the 20 improving school systems we studied show that strong commonalities exist in the nature of their journeys. This report outlines the aspects of these journeys that are universal, those that are context-specific, and how the two interact. We hope these experiences benefit school systems around the world in navigating their own path to improvement.

Standing on the Shoulders of Giants

An American Agenda for Education Reform

By

Marc S. Tucker

24 May 2011

National Center on Education and the Economy

Standing on the Shoulders of Giants: An American Agenda for Education Reform

by Marc S. Tucker

This paper is the answer to a question: What would the education policies and practices of the United States be if they were based on the policies and practices of the countries that now lead the world in student performance? It is adapted from the last two chapters of a book to be published in September 2011 by Harvard Education Press. Other chapters in that book describe the specific strategies pursued by Canada (focusing on Ontario), China (focusing on Shanghai), Finland, Japan and Singapore, all of which are far ahead of the United States. The research on these countries was performed by a team assembled by the National Center on Education and the Economy, at the request of the OECD.

A century ago, the United States was among the most eager benchmarkers in the world. We took the best ideas in steelmaking, industrial chemicals and many other fields from England and Germany and others and put them to work here on a scale that Europe could not match. At the same time, we were borrowing the best ideas in education, mainly from the Germans and the Scots. It was the period of the most rapid growth our economy had ever seen and it was the time in which we designed the education system that we still have today. It is fair to say that, in many important ways, we owe the current shape of our education system to industrial benchmarking.

But, after World War II, the United States appeared to reign supreme in both the industrial and education arenas and we evidently came to the conclusion that we had little to learn from anyone. As the years went by, one by one, country after country caught up to and then surpassed us in several industries and more or less across the board in pre-college education. And still we slept.

Until US Education Secretary Arne Duncan asked the OECD to produce a report on the strategies that other countries had used to outpace us, and then called an unprecedented meeting in New York City of education ministers and union heads from the countries that scored higher on the education league tables than the United States. Now, once again, the United States seems to be ready to learn from the leading countries.

In this paper, we stand on the shoulders of giants, asking what education policy might look like in the United States if it was based on the experience of our most successful competitors. We rely on research conducted by a team assembled by the National Center on Education and the Economy, at the request of the OECD, which examined the strategies employed by Canada (focusing on Ontario), China (focusing on Shanghai), Finland, Japan and Singapore. But we also rely on other research conducted by the OECD, by other researchers and, over two decades, by the National Center on Education and the Economy.

The policy agenda presented here is not a summary of what all the nations we studied do. There are few things that all of the most successful countries do. In the pages that follow, we will point out when all appear to share a policy framework, when most do and when some do. Companies that practice industrial benchmarking do not adopt innovations only when all of their best competitors practice them. They adopt them when the innovations of particular competitors appear to work well and when they make sense for the company doing the benchmarking in the context of their own goals and circumstances. Their hope is that, by combining the most successful innovations from individual competitors in a sensible, coherent way and adding a few of their own, they can not only match the competition, but improve on their performance. That is the approach we have taken here.

We contrast the strategies that appear to be driving the policy agendas of the most successful countries with the strategies that appear to be driving the current agenda for education reform in the United States. We conclude that the strategies driving the best performing systems are rarely found in the United States, and, conversely, that the education strategies now most popular in the United States are conspicuous by their absence in the countries with the most successful education systems.

Many will be quick to point to exceptions to our characterizations of American practice. In fact, examples of excellent practice in almost every arena of importance can be found in the United States. But our aim here is not to focus on isolated examples of good practice but rather on the policy *systems* that make for effective education *systems* at scale, for it is there that the United States comes up short.

We know that the complete transformation of the whole system of policy and practice we have suggested will seem an overwhelming prospect to many people. So we turn to Canada as our best example of a country that might be used as a source of strategies for making great improvements in the short term. It seems quite plausible that, while the short term plan is unfolding, the nation might embark on the longer term agenda we suggested earlier, which would lead to even greater improvements.

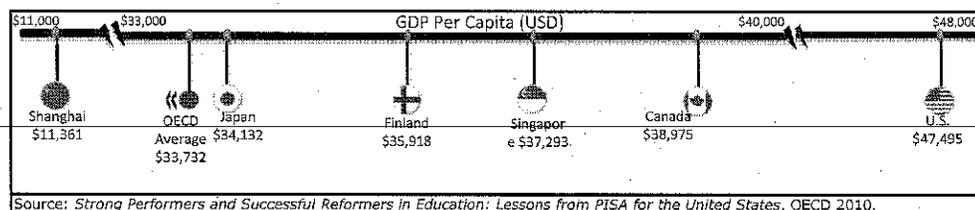
As you read this paper, bear in mind that, although we think there are useful roles that the United States government can play in improving dramatically the performance of our schools, we believe the main player has got to be state government. When we speak of changing the system, it is the states, not the national government, we have in mind.

So we begin by identifying broad themes, principles, policies and practices that appear to account for the success of some of the best-performing systems in the world.

The Broad Themes

Just below, we begin a detailed analysis of the strategies used by the countries with the most effective education systems. But it is easy to lose sight of the forest when looking at the trees.

The big story is about the convergence of two big developments. The first has to do with the trajectory of global economic development. The second has to do with the kinds of people needed to teach our children in the current stage of global economic development.



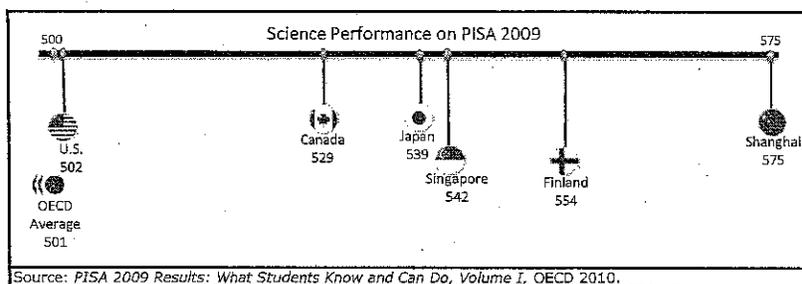
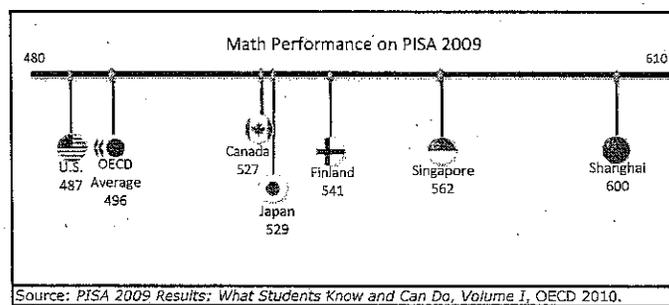
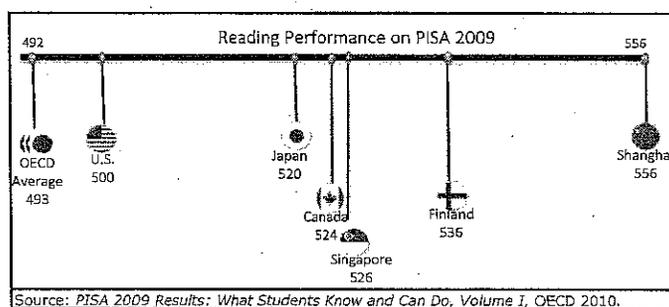
The nations we have described are either already very high wage countries or want to be very high wage countries. They have all recognized that it will be impossible to justify high relative wages for skills that are no greater than those offered by other people in other parts of the world who are willing to work for less, because we are all competing with each other now. Only those who can offer the world's highest skill levels and the world's most creative ideas will be able to justify the world's highest wages. These nations have also realized that this formulation means that very high wage nations must now abandon the idea that only a few of their citizens need to have high skills and creative capacities. This is a new idea in the world, the idea that all must have an education formerly reserved only for elites. It leads to abandonment of education systems designed to reach their goals by sorting students, by giving only some students intellectually demanding curricula, by recruiting only a few teachers who are themselves educated to high levels, and by directing funding toward the easiest to educate and denying it to those hardest to educate. It is this fundamental change in the goals of education that has been forcing an equally fundamental change in the design of national and provincial education systems.

The second big development follows from the first. No nation can move the vast majority of students to the levels of intellectual capacity and creativity now demanded on a national scale unless that nation is recruiting most of its teachers from the group of young people who are now typically going into the non-feminized professions. Recruiting from that pool requires a nation not just to offer competitive compensation but also to offer the same status in the society that the non-feminized occupations offer, the same quality of professional training and the same conditions of work in the workplace. Doing all that will change everything: the standards for entering teachers colleges, which institutions do the training, who is recruited, the nature of the training offered to teachers, the structure and the amount of their compensation, the way they are brought into the workforce, the structure of the profession itself, the nature of teachers' unions, the authority of teachers, the way they teach and much more.

Everything that follows is a gloss on the two preceding paragraphs. If they are right, if these are the core lessons from the countries that are outperforming the United States, then much of the current reform agenda in this country is irrelevant, a detour from the route we must follow if we are to match the performance of the best. We turn now to the details.

What the Top Performers Do And We Don't

We define a high-performing national education system as one in which students' achievement at the top is world class, the lowest performing students perform not much lower than their top-performing students, and the system produces these results at a cost well below the top spenders. In short, we said, we defined top performers as nations with education systems that are in the top ranks on quality, equity and productivity. In the following section, we summarize some of the key factors contributing to first-class performance in each of these three categories. We hasten to point out that this schema is rather artificial. System features described under any one of these three categories more often than not contribute to outcomes in others. System effects abound. Nonetheless, we think this schema will prove useful to the reader.



Before we get to the factors that most affect quality, equity and productivity, we point to the importance of international benchmarking as a key strategy for improving national education systems.

Benchmark the Best

Every one of the top performers is very conscious of what the other top performers are doing, though some benchmark more aggressively than others. The modern Japanese school system owes its very existence to trips taken by the new government when the Meiji Restoration took place, when the Japanese government resolved that the only way it could catch up with the West was to aggressively research its educational institutions and adopt and adapt the best of what they found. Japan has continued to research the education programs of the leading countries as a major input into its policymaking in education. The Singaporeans may be the most determined and disciplined benchmarkers in the world, not just in education, but across all fields of social policy. And their efforts have paid off. Finland has always made a point of researching the best performers when developing education policy. The current Premier in Ontario Province travelled abroad personally to visit other countries before settling on his new education policies for Ontario. The Hong Kong government actually hired an Australian who had done state-of-the-art work in several countries on curriculum, standards and assessment when they were looking for someone to reform their standards and assessment system.

Many Americans think that they have benchmarked other countries' education systems when they have established equivalency tables showing which scores on key American assessments correspond to certain scores on the national assessments used in other countries. But that is not what international benchmarking in education is for the countries that have been doing it for years. For those countries, to benchmark another country's education system is to compare broad goals, policies, practices and institutional structures as well as relative standing on common measures, in order to understand what another country is trying to achieve, how they have gone about achieving it, what they would have done differently if they could have done so, what mistakes they made and how they addressed them, which factors most account for their achievements and so on. Benchmarking is a wide-ranging research program that never ends, because no country's education system stands still very long.

Countries that base their education strategies on the careful study of successful strategies employed by the leading nations are not as likely to go down blind alleys wasting large amounts of resources on initiatives that fail to pay off as countries that base their strategies on untested theories, which is what the United States has tended to do over the years. What follows is a distillation of what the researchers affiliated with the National Center on Education and the Economy have learned since 1989 from the countries with the best education systems, with a particular focus on the countries, provinces and cities highlighted in this paper.

Design for Quality

Getting the Goals Clear

Reading the official documents from the ministries of the top-performing countries, and listening to the top officials in those countries, one cannot help but be struck by the attention that is being given to achieving clarity and consensus on the goals for education in those countries. It is probably no accident that Finland, Japan, Shanghai and

Singapore are without physical resources. All of these places have known for a very long time that their standard of living depends entirely on the knowledge and skills of their people. All now realize that high wages in the current global economy require not just superior knowledge of the subjects studied in school and the ability to apply that knowledge to problems of a sort they have not seen before (the sorts of things that PISA measures), but also a set of social skills, personal habits and dispositions and values that are essential to success. The Asian countries in particular are concerned that their students may not have as much capacity for independent thought, creativity and innovation as their countries will need. Though all these countries are concerned about developing the unprecedented levels of cognitive skills and non-cognitive skills required by the global economy, they are no less concerned about social cohesion, fairness, decency, tolerance, personal fulfillment and the transmission of the values that they feel define them as a nation. In many cases, these discussions of national goals have laid the base for sea changes in the design of national education systems, providing a solid foundation in national opinion for the kind of political leadership needed to redesign institutions that are—and should be—very hard to change. Not since the formation of the National Education Goals Panel in 1990, more than 20 years ago, has there been a focused discussion of America's goals for its students of the sort that many of these other countries have had more recently.

Instructional Systems and Gateways

Virtually all high-performing countries have a system of gateways marking the key transition points from basic education to upper secondary education, from upper secondary education to university, from basic education to job training and from job training into the workforce. At each of these major gateways, there is some form of external national assessment. Among the countries we studied, only Canada does not have such a system. Among the top ten countries in the PISA rankings, Canada is again the only outlier.

The national examinations at the end of upper secondary school are generally—but not always—the same examinations that the universities in that country use for entrance examinations. In many countries, these examinations are the only thing taken into account in determining who is admitted to which university and to the programs or schools within the university. It is also true, in many of these countries, that the scores on one's exams determine whether one will be admitted to upper secondary programs designed to prepare the student for admission to university. The content of the upper secondary exams is usually determined by the university authorities, and is closely tied to the content of the upper secondary curriculum. It is also typically true that there is an upper secondary program available to students who have successfully completed their basic education by the end of grade nine or ten that is intended to provide training for students who will either enter the job market when they complete it or go on to a polytechnic school for advanced technical training. The standards for the examinations at these gateways are typically set by the state in close collaboration with representatives of the industries that will employ the graduates, and, in some cases, with representatives of the labor organizations in those industries.

In the systems just described, there is very close alignment between the upper secondary curriculum, the upper secondary exams, and the university requirements. There is also very close alignment between employer's requirements and the skills students acquire to prepare for work in the industries in which they seek jobs. And finally, in these systems, regardless of which path a student decides to take in upper secondary education, they must all meet a common basic education standard aligned to a national or provincial curriculum before moving on to upper secondary school.

In countries with gateway exam systems of this sort, every student has a very strong incentive to take tough courses and work hard in school. Students who do not do that will not earn the credentials they need to achieve their dream, whether that dream is becoming a brain surgeon or an auto mechanic. Because the exams are scored externally, the student knows that the only way to move on is to meet the standard. Because they are national or provincial standards, the exams cannot be gamed. Because the exams are very high quality, they cannot be 'test prepped'; the only way to succeed on them is to actually master the material. Because the right parties were involved in creating the exams, students know that the credentials they earn will be honored. When their high schools say they are "college and career ready," colleges and employers will agree.

But the power of this system does not end there. In the countries that have some form of the system just described, the examinations are set to national standards and are directly derived from a national curriculum. Teachers in those countries are taught to teach that curriculum. It is also the case that these countries work out a curriculum framework, which means they decide, as a matter of policy, what topics should be taught at each grade level (or, in some cases, pair of grade levels) in each of the major subjects in the curriculum. In this way, they make sure that each year the students are taking the material that will be prerequisite to the study of the material that they are supposed to master the following year and all students will be ready for advanced material when it is offered. In these countries, the materials prepared by textbook publishers and the publishers of supplementary materials are aligned with the national curriculum framework.

Thus the standards are aligned with the curriculum, which is aligned with the instructional materials available to teachers. And the examinations are also aligned with the curriculum, as is the training that prospective teachers get in teacher training institutions.

In all of the countries studied for this paper, the national curriculum goes far beyond mathematics and the home language, covering, as well, the sciences, the social sciences, the arts and music, and, often, religion, morals or, in the case of Finland, philosophy. In most of these countries, few, if any, of the upper secondary school examinations are scored by computers and much of the examination is in the form of prompts requiring the student to work out complex problems or write short essays. They do this because the ministries in these countries have grave doubts about the ability of computers to properly assess the qualities they think most important in the education of their students.

Perhaps most important, the curricula and examinations in every country studied for this report, save Canada, were set not just to a very high standard, but to a particular kind of standard. Their students did well on the PISA examinations because they demonstrated high mastery of complex content as well as the ability to apply what they learned to practical problems of a kind they were not likely to have practiced on. Shanghai, Japan and Singapore have in recent years all engaged in multi-year massive revisions of their curricula to see if they could strike the right balance between high-level content mastery, problem-solving ability, and the ability to demonstrate a capacity for independent thought, creativity and innovation. Finland, having produced an elegant curriculum specification years ago for every level of their school system, has been making it less voluminous, in an effort to find the right balance between specificity and flexibility for their teachers.

The level of detail at which the national standards and curriculum are specified varies widely. In most of the East Asian countries, they are fairly detailed. In Finland, as just noted, they have been getting progressively briefer. In all cases they are guidelines, and in no case do they get down to the level of required lesson plans. They typically give teachers considerable latitude with respect to the specific materials used, pedagogy and pace.

It is important to point out that the United States has, in this realm, something that these other countries do not have, and it is not entirely clear that it is a good thing. The idea of grade-by-grade national testing has no takers in the top-performing countries. These countries do national testing at the gateways only, and some do not do state or national testing at every gateway. Typically, there are state or national tests only at the end of primary or lower secondary education, and at the end of upper secondary school. Schools and the teachers in them are expected to assess their students regularly as an indispensable aid to good teaching, but the assessments given between gateways are not used for accountability purposes, as the basis of teachers' compensation or to stream or track students.

Nonetheless, what has just been described is a very powerful instructional system that has few parallels in the United States. For a long time, Americans have preferred 'curriculum neutral' tests to those aligned with curriculum, virtually guaranteeing that students would be measured on a curriculum the teachers had not taught. Schools of education had no obligation to teach prospective teachers how to teach the national or state curriculum, because there was no such thing. Because the states had no curriculum frameworks, textbook manufacturers put a vast range of topics in their textbooks, knowing that any given topic might be taught by teachers at many different grade levels, and gave each of those topics only cursory treatment, because so many topics had to be included in the text. The federal government now requires tests in English and mathematics at many grade levels and has tied important consequences to student performance on those tests, thus heavily biasing the curriculum toward the teaching of these subjects and away from the teaching of other subjects in the curriculum that these other countries view as critical. Whereas these top-performing countries have placed a high value in their national policies on the mastery of complex skills and problem solving at a high level, the United

States has in recent years emphasized mastery of basic skills at the expense of mastery of more advanced skills. We continue to prefer tests that are largely based on multiple choice questions and that are administered by computers.

The new Common Core State Standards for mathematics and English and the work being done by the two assessment consortia will begin to address some of these issues, but, even when that work is done, the United States will still be at an enormous disadvantage relative to our competitors. We will have tests in these two subjects that are still not squarely based on clearly drawn curricula. The two consortia are betting heavily on the ability of computer-scored tests to measure the more complex skills and the creativity and capacity for innovation on which the future of our economy is likely to depend. No country that is currently out-performing the United States is doing that or is even considering doing that, because they are deeply skeptical that computer-scored tests or examinations can adequately measure the acquisition of the skills and knowledge they are most interested in. If the United States is right about this, we will wind up with a significant advantage over our competitors in the accuracy, timeliness and cost of scoring. If we are wrong, we will significantly hamper our capacity to measure the things we are most interested in measuring and will probably drive our curricula in directions we will ultimately regret.

In any case, if the interstate consortia continue to measure performance only in mathematics and English (with the eventual addition of science), we will have no multi-state curriculum and assessments in the other subjects in the curriculum for which many other countries have excellent assessments. It is unclear to what extent there will be strong curriculum and related instructional materials available to support the new tests in math and English, to say nothing of the other subjects in the broader core curriculum or subjects that cut across the curriculum. Nor is it clear to what extent our schools of education will assume responsibility for preparing teachers to teach the curriculum that emerges from the new Common Core State Standards efforts.

All of this is to take nothing away from the enormous achievement that is represented by the Common Core State Standards. But it is important to recognize that the development of the kind of complex, coherent and powerful instructional systems just described took many years to develop and improve in the countries we have studied. There is little doubt that these systems now constitute one of the most important reasons for their excellent performance. Implementation of the Common Core State Standards will still leave the United States far behind in what is undoubtedly one of the most important arenas of education reform. It will be essential to continue, to expand, and to expedite that work.

Teacher Quality

What we mean by 'teacher quality'

There is a good deal of discussion now about teacher quality, but it is not clear that there is much consensus as to what is meant by that term. But it is possible to derive a tripartite definition of teacher quality from the experience of the five countries we

studied: 1) a high level of general intelligence, 2) solid mastery of the subjects to be taught, and 3) demonstrated high aptitude for engaging students and helping them to understand what is being taught. We will take each in turn.

Some law firms in the United States recruit only from a handful of top universities. Others are happy to take graduates from the local night law school. The former firms recruit from the most elite universities not because they believe those universities do a better job of teaching the specific skills they are looking for but because they are using the university selection system to do their screening for them on some other qualities they care very much about. They are looking for people of outstanding general intelligence who also have the drive, tenacity and capacity for hard work that it takes to get into and survive the top law schools. They know that such people will quickly learn on the job what they need to know to do the specialized work they will be assigned. They know that, everything else being equal, they can count on such people to outperform their competitors on a wide range of assignments. They will be able to function with less supervision. They will produce better work. They will rise up the ladder of responsibility faster. The Japanese call this bundle of qualities "applied intelligence." Companies of all kinds in all industries will go as far up the applied intelligence scale as they think they can afford to secure a competitive advantage in their markets.

When a country is in the preindustrial stage or in the throes of a mass production economy, few workers will need advanced skills, and most students will not need much more than the basics. But, in advanced post-industrial economies, a much larger portion of the workforce needs to grasp the conceptual underpinnings of the subjects they study in school. They need more advanced knowledge. They need to be fluent at combining knowledge from many different fields to solve problems of a kind their teachers never anticipated. One can only do this with a much deeper and more advanced knowledge of the subjects in the core curriculum than used to be the case. And deep subject matter knowledge is not enough, either. They will have to be able to synthesize established and new knowledge quickly, analyze problems quickly and from odd angles and synthesize the knowledge they need in unusual ways to come up with creative and often unique solutions. They will need good taste as well. The students will not have that knowledge, those skills and the other attributes just mentioned if their teachers lack them. As we will see below, the top-performing countries are making strenuous efforts to greatly improve the subject matter knowledge of their teachers as well as their ability to analyze and synthesize what they know. So deep subject matter knowledge as well as the ability to use that knowledge effectively is the second requisite.

But one may be good at physics and still be a poor physics teacher. To be good at teaching, one has to be able to connect with students, to engage them, inspire them, communicate easily with them, get inside their heads and figure out what they don't understand and find a way to help them understand it. And it is not all about conveying 'content.' It is also about helping students to understand what the right thing is and why it is important to do it when doing it is not easy. It is about persuading a student that she has what it takes to go to college or stay in high school when her dad just went to jail and she is living on the sidewalk. It can be about being a friend, a mentor and a guide.

Most of the countries we studied have made strenuous efforts to raise the quality of their teachers in each one of these dimensions. The strategies they have used are sometimes very similar and sometimes very different.

Quality of the pool: Status, Compensation, Professional Working Conditions

Organizations that care about the quality of their workforce know that the single most important factor in that calculus is the character of the pool from which it recruits. No private firm, much less an entire industry, would prefer to recruit its professional staff from the least able college graduates if it could do better than that.

Three things directly affect the quality of the pool from which a nation recruits its teachers: 1) the status of teaching in the eyes of the potential recruit, relative to the status of other occupations to which he or she aspires, 2) the compensation offered, relative to other possible choices, and 3) the conditions of work, meaning the degree to which the way the work is organized makes it look more like professional work than blue-collar work.

It turns out that the countries with the most successful education systems are using a whole set of connected strategies to address all of these factors at the same time that they are addressing the need to get the teachers with the highest possible applied intelligence, the deepest content knowledge and the best teaching ability. Here's how they are doing that:

Standards for entry to teacher education

The logic for raising standards for getting into teacher education programs is the same everywhere. Low standards for entry means that people who could get into professional programs perceived as hard to get into see teaching as attractive only to people who do not have the skill or ability to do anything else, so they do not want any part of them. If these schools and programs are easy to get into, the message in the college or university is that they are low status and so higher education faculty who can get higher status jobs in their institutions do not want to teach in the education programs. Raising the standards for admission will attract a higher quality of applicant, and, at the same time, discourage lower quality applicants, and it will also attract a higher quality faculty, which also attracts a higher quality applicant.

So at this stage of the process, when applicants for teacher education programs are being considered for admission, quality means scores on common, highly regarded measures of general intelligence such as, in the United States, the ACT and the SAT; high scores or grades in courses in the subjects the applicant plans to teach; and high scores on relevant indicators that show the candidate has the personal attributes needed to connect with, inspire and support children of the ages he or she plans to teach.

We pointed out earlier that the Japanese have had high standards for entry into the teaching profession since the days of the Meiji Restoration more than a century ago.

Shanghai has raised their standards for entry into higher education programs intended to prepare teachers. Below, we describe how two other top performers go about making these determinations.

In Singapore, young people for a long time have taken “A Level” exams to get into teachers college. These are very difficult end-of-course examinations built on the English model. Low scores on these exams used to be sufficient for aspiring teachers, but, in recent years, that is no longer true and scores in the middle of the range are now required. Alternatively, the candidate can now present a polytechnic diploma, which is roughly equivalent to a high-level college degree in the United States. This is an even finer screen, because the polytechnics are in the top of the status hierarchy of the Singapore higher education system. In addition, the successful candidate must also survive a demanding interview conducted by a panel including National Institute of Education faculty, chaired by a serving or retired principal. The panel is charged to find out whether the candidate has the passion, commitment, communication skills, empathy and disposition to be a good teacher. Only one out of eight applicants survive this whole process.

In Finland, applicants for admission to teachers college who are accepted must survive a two-stage review. The first stage is a document review. To make it through this stage, they must: 1) score very high on the national college entrance exams, 2) have a high grade point average on their high school diploma and 3) have a strong record of out-of-school accomplishments while in high school. In the second phase they must: 4) complete a written exam on assigned books in pedagogy, 5) interact with others in situations designed to enable a skilled observer to assess their social interaction and communication skills, and 6) survive interviews in which they are asked, among other things, to explain why they have decided to become teachers. They are admitted to a teacher education program only after they have passed all of these screens. Only one out of ten applicants for entry into Finnish teachers colleges are admitted.

Thus two of the countries with the highest scores on the 2009 PISA have both instituted rigorous measures used to determine entrance into teacher preparation programs intended to assess all three of the components used to define teacher quality at the beginning of this section. The effect of these rigorous measures is to limit Singapore’s intake to the top 30 percent of high school graduates and to limit Finland’s intake to the top 20 percent.

It is a different story in the United States. The College Board reported in 2008 that when high school graduates going on to college were asked what their intended major was, those who had decided on education scored in the bottom third on their SATs. Their combined scores in mathematics and reading came in at 57 points below the national average.

This should not surprise us, because, in our country, most schools of education at both the undergraduate and graduate levels are widely regarded as very easy to get into. Their status within the university is typically among the lowest of all schools and departments.

This was often the case in the best-performing countries not so long ago, before they began their march to their present much higher rankings.

There is, of course, a shining exception to this broad generalization, which is Teach for America, which famously enrolls very high-performing graduates of many of the most elite colleges in the United States and then assigns them to teaching positions in schools serving disadvantaged students. But Teach for America only serves to underscore the point being made here. The proportion of openings for new teachers every year in the United States filled by Teach for America participants is vanishingly small, and, in any case, most have no interest in continuing as career teachers after they have satisfied the initial requirement anyway. Teaching is viewed by many Teach for America participants as the equivalent of a tour in the Peace Corps, not as a serious career opportunity. The experience of Teach for America makes it plain that it is possible to attract the very best and brightest to teaching, but Teach for America does not itself provide a path to staffing our schools with highly capable teachers for the time and in the numbers needed. Teach for America is not an alternative to building schools of education that can attract first rate candidates and teach prospective teachers what they will need to know to be successful in our schools.

It has not always been this way. There is reason to believe that the standards for admission to teacher education programs in the United States were once considerably higher.

In fact, there is reason to believe that the problem with the American teaching force is not that it has long been of low quality and must now be raised, but rather that the United States greatly benefitted for the better part of a century from having a teaching force largely made up of college-educated women whose choice of career was largely limited to nursing, secretarial work and teaching, and some minorities whose career choices were similarly constrained. Many women chose teaching because it would allow them to be home when their children came home from school. Because career choices were so limited, the American public reaped the twin blessings of a highly capable teaching force willing to work for below-market wages under poor working conditions. Those who accepted that deal are now leaving the workforce in droves. There are now more women than men in the professional schools preparing young people for many of the most prestigious professions and they are taking advantage of those opportunities. The United States is now about to get the least capable candidates applying to our education schools when we need the best.

When we had a higher quality candidate applying to our teachers colleges, the colleges could afford to be more selective. That is why there is good reason to believe that the standards for entry into teacher education have been sliding. When the baby boom was leaving our colleges, many people predicted that the coming baby dearth was going to result in great reductions in the size of college student bodies as the size of the whole cohort declined massively. But, though the size of the cohort certainly declined, the size of student bodies did not. The data suggest that the colleges made a fateful decision to lower their standards to fill their classrooms. There is every reason to believe that this

happened in our teachers colleges in just the same way it happened in other colleges, but it was also at this time that opportunities for women and minorities greatly expanded, which would mean that the quality of applicants in teachers colleges would have suffered from both of these causes, not just one. Furthermore, analysts are now noticing a large falloff in applications for admission to teachers colleges all over the country, a result of the financial crisis. Potential candidates, who used to view teaching as almost immune from the business cycle and therefore one of the most secure of all occupations, are noticing that teachers are being laid-off in very large numbers and now see teaching as a very risky bet.

Put these three points together—highly qualified college educated women and minorities abandoning teaching as a career, the drop in admission standards following the baby boom and the decision by many capable students to avoid teaching because of the widespread teacher layoffs, and we can see the danger ahead for the United States. All we need to do to acquire a very poor teaching force is nothing. Inaction, not action, will bring about this result. It is critical that this trend be reversed. We cannot afford to continue bottom fishing for prospective teachers while the best performing countries are cream skimming.

Attracting top flight students to teacher education and a career in education—the compensation angle

Most of our competitors have formal policies that peg teachers' compensation to the top ranges of their civil servants' compensation system or to the compensation of other professionals, such as engineers, in the private sector. Their aim is to make sure that young people making career choices see teaching as offering compensation comparable to that offered by the more attractive professions. Finland's teachers appear to get paid a little less, relatively speaking, than teachers in the other top countries, but, because salaries for everyone are very flat in Finland compared to most other countries, and the status of teachers is so high, they still get excellent candidates.

At the International Summit on the Teaching Profession convened by Secretary Duncan in New York City in March 2011, the Minister of Education of Singapore offered the observation that the goal of compensation policy ought to be to “take compensation off the table” as a consideration when able young people are making career decisions. There was wide agreement on that point among the ministers of the other top-performing countries around the table.

The United States is far from the Singapore minister's standard. According to the National Association of Colleges and Employers, teachers earn a national average starting salary of \$30,377. That compares with \$43,635 for computer programmers, \$44,668 for accountants and \$45,570 for registered nurses. None of these occupations are among the leading professions, which provide starting salaries that are even higher. Not only do teachers make markedly less than other occupations requiring the same level of education, but census data shows that teachers have been falling farther and farther behind the average compensation for occupations requiring a college degree for 60 years.

The average earnings for workers with college degrees are now 50 percent higher than average teachers' salaries, which is a very long way indeed from the Singapore minister's standard.

Making sure that initial and average compensation for teachers is competitive is essential. But there are other issues having to do with compensation and financial incentives for choosing teaching as a career that other nations have addressed and we have not.

Shanghai, for example, has waived its charges for tuition for teacher education and offered early admissions to students applying to teacher education programs. This has made teaching a very attractive career choice, especially for students from the poorer provinces with strong academic backgrounds. Though the compensation for teachers in China is low by international standards, teachers in that country can make substantial additional income from tutoring. And the government also offers bonuses to teachers willing to teach in rural areas. The result of these and other initiatives has now made teaching the second or third most popular career choice in China, a very recent development.

It is obvious on the face of it that if compensation is not adequate, raising standards for admission to teacher preparation programs in universities, raising the standards for licensure and refusing to waive those standards in the face of teachers' shortages will simply guarantee shortages of teachers into the indefinite future.

It turns out that total compensation of teachers is more competitive than cash compensation taken by itself, because American teachers' compensation, like that of civil servants generally, is heavily weighted toward retirement benefits. Costrell and Podgursky report that, in 2008, employer contributions to teachers' retirement plans was 14.6% of earnings, compared to 10.4% for private professionals, this difference having more than doubled in the four years since the data were first collected. The problem with this is that, while it provides a strong incentive for experienced teachers to stay in teaching longer than they might otherwise, it makes teaching unattractive to young people who are more concerned about supporting new families than about their retirement.

The trajectory of cash compensation is also important. Most American teachers top out quickly. And, even when there are adjustments for differences in the quality of teaching, which is very rarely done, they are very small. Countries that are restructuring teachers' careers are adjusting compensation as teachers ascend career ladders within the profession and in administration, and take on more authority and responsibility as they do so. We have also seen that some countries—again, Singapore is a good example—are paying bonuses of up to 30 percent to teachers who are found to be particularly effective on a wide range of measures. And many of those countries, not just China, are paying more to teachers who are willing to work in outlying areas or who bring qualifications in short supply.

Institutional setting

As late as the 1970s, Finnish teachers were prepared in relatively low status colleges dedicated to teacher education. Now, all their teachers are educated in their major universities. This was not accomplished by simply allowing the former teachers colleges to become universities, but by sending prospective teachers to institutions with the highest status in the postsecondary system.

Years ago, prospective teachers in Singapore were also trained in a separate and relatively low status college for teaching. Then, Singapore created the National Institute of Education to train its teachers. More recently, the government incorporated the NIE into Nanyang Technological University, a top tier institution in Singapore's higher education system. Nanyang has partnerships with many of the world's most highly regarded research universities and is ranked by *The Economist* as having one of the best business schools in the world. NIE is now a major research institution in its own right, and, at the same time, a very high status part of Singapore's postsecondary education system.

Thus many of these top-performing countries have not only greatly raised their standards for getting into higher education institutions preparing teachers, but most have moved teacher education out of their lower tier institutions and into their top tier institutions. This has had the effect of further raising the status of teaching, improving the quality of faculty, improving the quality of research on education, facilitating the dissemination of high quality research to prospective teachers and creating a teaching force that is less likely to put up with old forms of work organization once they become school teachers.

Teacher education in the United States is no longer done in institutions called normal schools, but it is generally done in second and third tier, relatively low status institutions, many of which were formerly normal schools. When it is done within major universities, it is typically accorded the low status associated with the other feminized occupations. While graduate education in education is often done in the major research universities, many of the institutions that offer professional training in school administration and education research do not offer professional training to school teachers. This is very similar to the profile that many of the leading countries abandoned ten or more years ago.

Content of teacher education and induction

We combine here two functions usually thought of quite separately: what prospective teachers are taught about their craft before entering service and what they are taught immediately after entering service. The reason we have done that is that some top-performing countries rely heavily on pre-service teacher education to teach the skills of the craft to teachers and some put much more emphasis on the use of apprenticeship-style instruction in the workplace to convey the essential craft skills, once the teacher has been hired by the schools. This is an important difference.

Consider first the approach taken by Finland. The Finns, as we have seen, require all of their teachers, including their primary school teachers, to have a master's degree. Primary

teachers major in education, but they must minor in at least two of the subjects in the primary curriculum. These minors are taken not in the education schools but in the arts and sciences departments of the university. Upper grade teachers must major in the subject they will be teaching. Their education in pedagogy is either integrated into their five-year program or provided full time in the master's year after the student has completed a bachelors program with a major in the subject that person will teach. Candidates who already have a master's degree in the subject they will teach must get another master's degree in teaching. There are no "alternative routes" to entering the teaching force in Finland. The only way to become a teacher in Finland is to get a university degree in teaching.

Clearly the Finns place a very high value on having teachers who have really mastered the subjects they will teach, and have also placed a high value on giving teachers the skills they will need to teach those subjects well once they arrive in the classroom.

Now consider the approach taken by Shanghai. In Shanghai, 90 percent of the teacher preparation program is devoted to mastery of the subject the prospective teacher will be teaching. A school mathematics teacher in training is expected to take the same undergraduate mathematics curriculum as undergraduates who will go on to do graduate work in mathematics, a very demanding curriculum.

It is clear that the Shanghai authorities are at least as determined as the Finns that the teachers who go on to teach science or any other subject know as much about the content of those subjects by the time they complete their undergraduate program as the people who will go on to be physicists or chemists or mathematicians know about those subjects when they complete their undergraduate program. And that is just as true of their future elementary school teachers as it is of their secondary school teachers.

The comparison with American policy and practice on the same point is very telling. Whereas elementary school teachers in these two other countries specialize in math and science or in social studies and language, Americans preparing to become elementary school teachers do not. Most American elementary school teachers know little math or science and many are very uncomfortable with these subjects. That is hardly true of their counterparts in Finland or Shanghai. And some of our secondary school teachers of math and science know a good deal less than their counterparts in those countries. It is also true that once one becomes a teacher in the United States, irrespective of the arena in which one is trained, a teacher can be assigned to teach a subject in which he or she was never really trained at all.

An anecdote related to this point is worth telling. Some years ago, Bill Schmidt, among the most distinguished of Americans who have been benchmarking the performance of the leading nations over the years, and who led the American team working on the TIMSS studies, was in a meeting with his other colleagues from the countries designing the tests and research studies. One of the Americans made a pitch for including a background question in the research instrument that would have asked how many teachers of mathematics and science in each country were teaching subjects they had not

been prepared to teach. There was an expression of astonishment from the representatives of all the countries except those from the United States. It simply was not done. Teachers were not permitted to teach outside their subject. There was no need to ask this question. The topic was never raised again. Evidently, only the United States, among all the industrialized countries, allows its teachers to teach subjects they have not been highly trained in.

The cumulative result of these differences is a much greater likelihood that, from the first grade to the last, school children in Shanghai and Finland are likely to be taught by teachers who have a better command of the subjects they will be teaching. The consequences of these differences are incalculable.

We come next to the question of policy and practice concerning the standards to be met by teachers in Shanghai, Finland and the United States with respect to pedagogy. It turns out that this is a very important issue in both Finland and Shanghai, but the strategies for achieving excellence in this important arena are very different in these two countries.

The Finns place most of their faith in developing the pedagogical skills of their future teachers while they are still in pre-service training. Obviously, the Finns believe it is very important for prospective teachers to get a strong background in pedagogy before entering the teaching force. They provide a strong background in the research underlying teaching and they also provide their teachers with strong research skills. All teacher candidates have to complete a research-based dissertation. Prospective teachers are expected to learn a lot about subject-specific pedagogy. There is considerable emphasis in the teacher education curriculum on the development of the candidate's skills at diagnosing student problems and learning how to choose the right solution for those problems, based on the relevant research. And there is a very strong clinical element in the program, including a full year of practice teaching done under the close supervision of a master teacher.

Whereas the Finns take five years or more to educate a teacher and divide that time almost equally between content training and pedagogical training, the Chinese, as we just saw, devote 90 percent of the available time during pre-service training to deep mastery of the subject the prospective teacher is preparing to teach. The remaining time available for teacher education—only 10 percent of the total—in Shanghai is devoted to a program of instruction in education theory, the psychology of learning and teaching methods that has not changed in many years and which many observers think is very out of date.

At first glance, that would appear to suggest that the Finns believe in the importance of substantial instruction in pedagogy and the Chinese think it unnecessary. But that is not the case. In Shanghai, a new teacher is expected to spend the first year of employment as a teacher under the intense supervision of a master teacher. Their master teachers are relieved of all or most of their classroom responsibilities to allow them to play this role. These master teachers often sit in on every lesson taught by the new teacher, providing intense coaching. And the new teacher will also observe the master teaching many lessons, too.

Recall that the Finns have decided that it is essential that their prospective teachers learn as much as possible about how to diagnose the nuances of the difficulties students experience in mastering difficult material, as well as how to identify the right techniques and methods to address those problems. The Finns put a lot of effort into building their prospective teachers' skills in this arena before they enter service. The Chinese are no less concerned than the Finns that their teachers master the art of teaching, but they have a very different strategy for accomplishing this aim. They put most of their faith in a very demanding apprenticeship strategy, as soon as the teachers-college graduates are hired.

Both countries devote a lot of resources to the development of the pedagogical skills of their recruits. It is certainly true that American schools of education teach methods courses. But American teachers complain constantly that what they learn in these courses is of very little value when they enter real classrooms. By all accounts, the efforts of the Finns and the Chinese to give their prospective teachers and beginning teachers much better supported initial classroom experience, at the hands of master teachers who can build their skills at recognizing specific problems that students have in learning the subjects they will teach and figuring out which research-based techniques are appropriate to address those problems, is an important key to those countries' success.

The careful attention to the development of skills in diagnosis and prescription, in the development of effective lessons, in the adjustment of instruction to the actual needs of students, under the extended and intensive guidance of master teachers, has no counterpart in the American experience. Little attention is typically devoted to detailed instruction in diagnosis and prescription, except, in some instances in the case of special education. The typical clinical experience of American candidate teachers is usually of poor quality, too brief, unconnected to the rest of their instructional program and provided by classroom teachers who cannot on the whole reasonably be called 'master teachers.' Once graduated from teachers colleges and hired by their first school district, they are typically put in a sink or swim situation, with little or no support from experienced teachers or supervisors, often in the most demanding classroom situations. Once again, the contrast with the experience of their Shanghai and Finnish colleagues could not be more stark.

It is worth pointing out, however, that the training of American medical doctors rests firmly on the very elements just described as the basis of the training of Finnish and Shanghai educators. American medical doctors are supposed to have a thorough background in the sciences that underlie medicine, physiology and pathology. Their training is essentially clinical in nature and is provided by master practitioners. The heart of the training is a form of apprenticeship known as rounds and residency. The most important aspect of their training is skills in diagnosis and prescription, based on a firm knowledge of the relevant research. This training takes place not in third tier, low status institutions, but in professional schools in top research universities. Most of these features have been adapted to the needs of professional education in teaching by most of our top competitors. None yet typify American practice.

Licensure and standards for entry

When teachers' shortages develop in the United States, the government's response is almost always to waive the regulations defining the minimum qualifications for teaching in public schools. When there is a shortage of civil engineers, we do not say that it is no longer necessary to make sure that civil engineers have the qualifications needed to design safe bridges nor, in such situations, do we decide that doctors no longer need to meet the minimum requirements for licensure. If there is a shortage in those fields, or indeed in virtually all truly professional fields, compensation increases until the market clears and the shortage disappears. There is no clearer sign of society's lack of respect for teachers and teaching than its view that, in the end, what really matters is having a warm body in front of their children, irrespective of that person's qualifications to teach. The best performing nations do not do this. They do not have to. They have, as we have seen, many more fully qualified applicants for teaching positions than positions available.

Continuing professional development: the instruction connection

It would not ordinarily come naturally to most Americans to combine these two topics, but that may be part of our problem, because it would come quite naturally to educators in many of the top-performing Asian countries.

Consider the Japanese practice of lesson study. In Japanese schools, the faculty work together to develop new courses or redesign existing courses to make them more engaging. Once developed, that course is demonstrated by one of the teachers and critiqued by the others and revised until the faculty is happy with it. Then a particularly capable teacher will demonstrate it for others and critique their practice when they in turn teach it. Throughout, the development process calls on the latest research. Teachers who get very good at leading this work are often called on to demonstrate their lessons to other schools and even to teachers in other districts and provinces. In this way, instructional development and professional development are merged and professional development becomes an integral part of the process of improving instruction in the school, informed by the latest and best research.

In fact, Japanese teachers are provided with research skills in their pre-service training, so that this local, teacher-led development process is supported by the kind of research skills needed by teachers to make sophisticated judgments about the effectiveness of their local development work.

In the United States, teachers are generally the objects of research rather than participants in the research process itself. The topics for professional development are often chosen by administrators in the central office rather than by teachers seeking to improve their own practice on terms of their choosing. Because the topics chosen for professional development are typically not the topics the teachers would have chosen, they often perceive the professional development they get as not particularly helpful. The Japanese model just summarized is certainly not the only model used in the top-performing

countries, but it suggests the possibilities that come to mind when teachers are viewed as highly competent professionals who are expected to take the lead in defining what good practice is, advancing that practice and keeping up to date on the latest advancements, which is exactly what happens in the professions that are led by the members of the profession rather than those who are administratively responsible for their work.

Continuing professional development: the career development connection

All over the world, well run companies and government agencies give a lot of thought, not only to how they can source their staff from the most capable pool possible, but also how they can offer their best people attractive careers in the agency or company, careers of increasing responsibility and authority, and the increased compensation and status that come with those jobs. Typically, they carefully groom their most promising staff for the next steps on the ladder, giving them at each stage the training they will need for the next job, providing mentors who can help them develop the right skills and so on.

That is what Singapore does for its professional educators. Having done their best to make sure that they have a very talented pool from which to source their teachers, they recruit the best and then provide top-level training for them. But it does not end there. They have carefully structured several distinct career lines that are available to the new recruits. For each career line, they have designed programs of training that are matched, step by step, to each step of the ladder. The system selects those people for further training who have the best qualifications, get the best ratings and have done the best in the training for the next position. In this way, Singapore carefully nurtures its talent pool, reserving the most expensive training for the people best prepared to use it well.

When teachers in Singapore are first hired, they become eligible to choose among three possible career ladders. One leads to the position of Principal Master Teacher through the intermediate steps of Senior Teacher, Lead Teacher, Master Teacher and, finally, Principal Master Teacher. That is the "Teaching Track." Teachers who want a career in administration proceed through Subject Head/Level Head, Head of Department, Vice Principal, Principal, Cluster Superintendent, Deputy Director, Director, and lastly, Director-General of Education, the top spot. That is the "Leadership Track". And there is another track, called the "Senior Specialist Track," designed to describe the trajectory of a career in the Ministry of Education in various specialized areas such as curriculum and instructional design, and education research and statistics. Highly qualified candidates for advancement in this system may be offered scholarships for advanced study in Singapore and abroad, in leading universities all over the world. They may be deliberately rotated among carefully selected assignments in the schools in the Ministry to give them the kind of experience the Ministry is looking for.

It is fair to say that neither the United States nor the individual states have policies designed to create a high quality pool from which we select candidates for teacher training. We often take whoever shows up. The pool is self-selected. With rare exceptions, we do not have well-defined career paths for teachers who want to advance their careers, but stay in teaching. Nor, obviously, have we defined the training and

further education that candidates for advancement on that nonexistent path must complete to be eligible for advancement. Indeed, we have not defined, as the Singapore government has, what qualities we are looking for in teachers that would qualify them for advancement.

This section on teacher quality is one of the longest in this paper and it is easy to lose the thread. But there is one.

We see two images, one of teaching in the United States and the other of teaching in the countries with the world's most effective education systems. They are very different.

As we have seen, the prevailing view in the United States is that our teachers need not come from the more able strata of the college-educated population. We behave as if we believe that only a few weeks of training is needed to do what they have to do, a sure sign that we do not believe teaching is a profession at all. If they do get more, it can certainly be done in very low-status institutions, and if they do not have much training, it is no big deal. If there is a shortage of teachers, we quickly waive the very low standards we insist on in boom times. We congratulate ourselves on offering \$10,000 signing bonuses to teachers when we worry about the qualifications of the ones we are getting, and then wonder why it does little to attract a better quality of candidate or simply more candidates. We do little or nothing about starting salaries that will not permit a young teacher to support a small family in the style to which college graduates are accustomed in this country. In most places, teaching continues to be a dead-end career, with no routes up except those that lead out of teaching. We make teachers the objects of research rather than the people who do research. We talk a lot about getting rid of the worst teachers, as if that was our biggest problem, but nothing about doing what is necessary to get better ones, thus accomplishing little but the destruction of teacher morale. We do all of this while talking a lot about teacher quality.

So it should surprise no one that we have a teacher quality problem.

When we looked at the countries topping the education league tables, we saw that teaching is not just referred to as a profession but is actually treated as though it is one. Those countries are willing to compensate teachers in the same way they compensate people in the professions, which, until recently, have been heavily dominated by men. They take their professional training seriously. It is lengthy and done in high prestige institutions. The standards for getting into those institutions are very high, and the competition to get into them on the part of top-notch students is quite stiff. The program of training mimics the way doctors and other highly regarded professionals are trained. They are carefully mentored by very capable people when they are hired. They are at the heart of the process of improving the system, not the object of that process, and their career prospects depend on their professional contribution, just as is the case for real professionals everywhere else. It would appear that the top-performing countries are far along in a process of converting their teachers from blue-collar workers to professionals

on a par with the other professions. Is it any wonder that these countries are experiencing much better results than the United States?

Of course, if teaching moves away from a Tayloristic work organization and takes on the attributes of a true profession, that will have implications for our teachers' unions and their contracts. American labor law is firmly grounded in the mass production model of work organization and assumes that workers and management will be locked in eternal conflict. The Taft-Hartley Act assumes that conflictual relationship and sets out the rules under which it will work. But, although that act of Congress was intended to apply only to the private sector, it was eventually applied to the public sector by most states, and that resulted in the work rules and contract provisions that are now giving this country so much trouble. Those rules can and should be changed. As the states decide to pay teachers like professionals and provide teachers the kind of professional responsibility and autonomy that other professions have, the teachers will need to be willing to write contracts that move away from the blue-collar model and toward contracts that embrace a professional model of work organization, in which teachers take responsibility for raising teaching standards to world-class levels, for the performance of students, for working as many hours as it takes to get the work done, for evaluating the work of their colleagues, recommending termination for teachers who do not measure up to high standards and so on.

Teachers will have to give up seniority rights of assignment and retention and other hallmarks of the blue-collar work environment and they will have to accept the proposition that some teachers will be paid more than others and have different responsibilities in recognition of their superior performance. That is part of what it means to be a professional. In exchange, of course, they will earn once again the high regard of the public and their peers, be paid like engineers and architects and doctors and enjoy the same high status in the community and their country that their colleagues in the top-performing countries enjoy.

Principal Quality

In much of the rest of the industrialized world, school leaders are called head teachers, because they continue to teach while they manage. Typically appointed because of their superior teaching ability, they are still viewed as teachers, but with additional responsibilities.

This is probably because schools in most other countries are smaller than American schools, but also because, in the United States, schools typically have less discretion, especially in the suburbs and cities, than in other countries, reporting to district central offices that are larger, often much larger, than their counterparts in most other countries. Having an intermediate layer of administration that is both larger and closer than it is elsewhere produces much more detailed and frequent requests and demands for information and compliance than school heads in most other countries experience. That, too, makes school leadership a full time job.

One result of this difference is that few of the countries with the most successful education systems have separate licensure for school heads or specialized training for them, though that is beginning to change, as many of the leading countries are now realizing that they may be able to improve their systems even further by attending more than they have in the past to the selection, training and licensure of school heads.

Singapore, an exception to the general rule, takes the training of school principals very seriously, offering, as we have just seen, a separate, defined career path for teachers who seek school and district leadership positions. Candidates for principal positions must take a six-month training program consisting of course work, supervised practice and mentorship, all monitored against clear definitions of the qualities that the Singapore government is looking for in their principals. The mentoring component of the program takes place during two sessions, each one a month long. Aspiring principals shadow principals hand-picked by the Ministry for their outstanding leadership qualities. The process is mediated by a faculty member from the National Institute of Education.

Instruction

The Japanese use an approach to instruction that can reasonably be described as whole class instruction or large group instruction but is definitely not lecturing. The teacher sets an assignment for the class, walks up and down the rows of students working the problem, picks out students using very different strategies for solving the problem, and asks the students who devised those strategies to come to the board—one by one—and describe their approach to the problem. The aim is not to focus on the right solution, but to provoke an extended class discussion of the various strategies used to get to a solution. This discussion of the strategies employed by the students is intended to help them understand *why* the right solution works, that is, to get to a deeper understanding of the topic under study than the American student typically gets by focusing only on the one method the teacher has decided to use to solve the problem. Because this technique depends for its success on identifying a good variety of solution strategies, teachers in Japan want large class sizes, not small ones. This approach to instruction is characteristic not only of Japan, but of many other East Asian countries as well.

Focusing on the relative effectiveness of different instructional strategies is obviously important in its own right, but it is also important because of the effects on other factors affecting student achievement.

Of all the strategies available to improve student performance, decreasing class size is among the most expensive and least effective. Instructional strategies that improve the outcome by increasing class size can release very large sums of money that can also improve student achievement, thus creating a very large multiplier effect. We will return to this point below in the discussion of tradeoffs in education system design.

But we should also note that the instructional methods used in Finland are different from those used in Japan, especially at the high school level. Though the Japanese are putting a relatively new emphasis on learning as distinguished from teaching, that is, on

promoting more student initiative in the learning process, Japanese teachers are still expected to stay pretty close to the national curriculum as promulgated by the Ministry, and that curriculum is pretty clearly spelled out. Finland, on the other hand, has been pressing hard in recent years toward a teaching and learning style in which the student takes increasing responsibility for the learning process. The Finns have been paring down the length of their curriculum guidance, and providing many more choices with respect to what is studied by modularizing the curriculum at the upper secondary level and letting the students assemble their own curriculum. This trend in curriculum is accompanied by a complementary trend in learning and instructional style, away from whole group instruction and toward problem- and project-based learning that is pursued individually and in teams. To the extent that students select and design their own projects and decide how to go about addressing them, this becomes student-directed learning in which the teacher becomes a facilitator rather than director of the learning process, and the object of instruction becomes not only the acquisition of subject-based knowledge and skill, but also the ability to frame problems to make them more amenable to solution, to identify possible sources of information that bear on the problem at hand, to analyze that information, synthesize what has been learned to frame a solution and then communicate the solution. What has just been defined is a disciplined learning process intended to enable the learner to come up with sophisticated and creative solutions to novel problems. Increasingly, this is the object of Finnish education. It requires teachers whose great skill is not so much the development of great lessons as teachers who are great stimulators, facilitators, mentors and partners in the learning process and who can create learning environments that are more like workshops than classrooms, whose intellectual skills and knowledge are deep enough and flexible enough for them to follow and lead their students in very unpredictable directions.

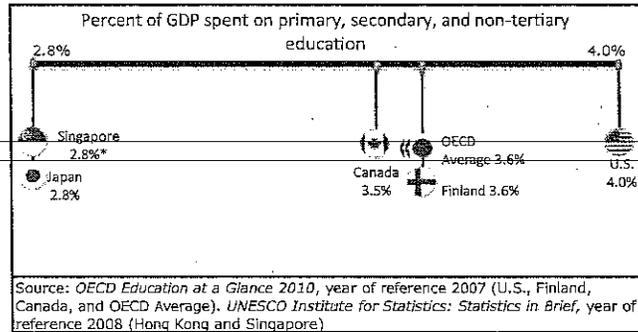
But we hasten to add that self-directed problem- and project-based learning can easily turn into a poor substitute for deep mastery of the underlying subjects in the curriculum. When the student lacks a firm command of the nuances of the core subjects in the curriculum, project- and problem-based curricula often result in very shallow knowledge gained in the classroom. What makes it work in Finland is the fact that these pedagogies and learning methods rest on top of solid mastery of the core subjects in the curriculum, acquired by Finnish students in the lower grades.

Design for Equity

School Finance

Local control of school finance has been an emblem of American education for a very long time, and is a deeply ingrained feature of our system. In essence, in many states, groups of citizens have been allowed to gather together to form their own education taxing districts. The result is that wealthy parents, by forming their own taxing districts, can drive their tax rates very low while benefitting from very high tax yields. At the other end of this spectrum, people who cannot afford very much for housing end up congregated together in districts where they must tax themselves at very high rates to produce a very low yield. In such a system, the children of the wealthiest families get the

best teachers and the best of all the other available education resources, and the families with the least money get the worst teachers and the worst of everything else as well.

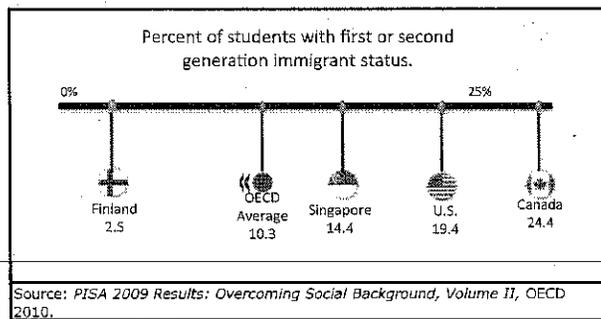


Almost all of the top-performing countries have been moving away from local control, if they ever embraced it, and toward systems designed to distribute resources in ways intended to enable all students to achieve high standards. That does not mean equal funding for all students; it means differential funding; it means unequal funding designed to come as close as possible to assuring high achievement across the board.

Perhaps the most interesting case from an American perspective is Canada. Two decades ago and more, elementary and secondary education in most of the provinces was funded much the way it is funded in the United States, with each locality raising much of the money locally, with the provinces providing additional sums intended to moderate the disparities in per student funding that such a system inevitably produces. But, about 20 years ago, this began to change. Conservative governments, in response to complaints from citizens about skyrocketing local tax rates, initiated a move to steadily reduce reliance on local taxes and to increase the portion of the total budget paid for by the province. In the biggest provinces now, little if any of the money for public education is raised locally. All or almost all comes from the province. Not surprisingly, the gross inequities that came with raising money locally are gone, too, and Canada, like the top-performing countries elsewhere, is moving toward a funding system intended to promote high achievement among all students, which means putting more money behind hard to educate children than children who are easier to educate.

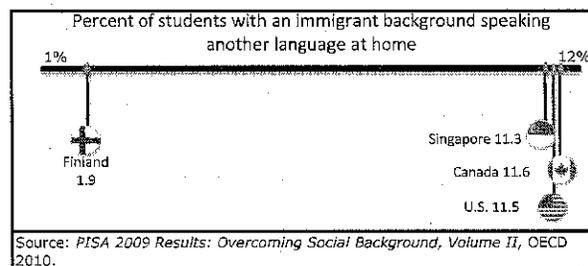
Secondary School Organization

When one looks far enough back in the history of most industrial nations, one usually gets to a time when their primary schools were comprehensive (in the sense that students from all social classes were mixed together in all or almost all the classrooms) and the upper grades were not. As secondary education developed in most countries, separate schools were created for three groups of students: the children of the working class, the children of the artisans and shopkeepers and the children of the nobles, or, later, the professionals, owners and managers of the larger enterprises.



In some countries, secondary schools were comprehensive in their enrollment, but, as in the United States, there were different tracks or streams within those comprehensive schools for the children of different social classes, so the result for the students was the same as in those countries that had different schools for students from different social classes. Depending on the country, the break between the comprehensive lower schools and the tracked upper grades might come as early as the end of grade four.

In the Scandinavian countries, after World War II, the period of comprehensive basic education for all students was extended to the point that most of the Scandinavian countries now have common schools through grades nine or ten. Students from all backgrounds attend these schools and they get the same curriculum. In these and some other countries, it is not until a student is sixteen that education paths begin to diverge.



Inevitably, as the previously separate education programs are merged and the decision to give all students substantially the same education is made, there is a national discussion about the standard to which that education will be set. In the countries with the high-performing education systems, that argument was almost always settled by a decision that the standard to be adopted would be the standard that formerly applied only to the students in the top track.

This battle took place in Japan more than a century ago and in Finland after the Second World War. Singapore abandoned streaming in its primary schools, but the standard for its lowest stream just above primary school is still well above the average standard of performance for the OECD nations. The United States calls its high schools comprehensive schools, but it still offers different courses set to very different challenge levels to students from different social backgrounds in most communities. The implementation of the Common Core State Standards might change that, but, for now, few American high schools expect most of their students to reach a global standard of

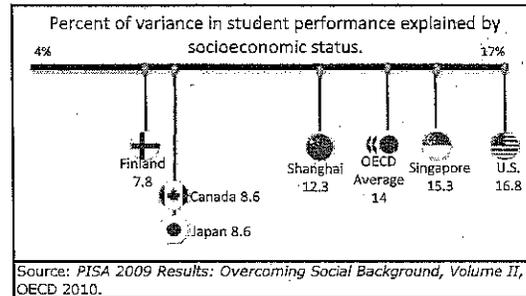
academic achievement by the end of grade nine or ten, though that is exactly what the top-performing countries are doing now.

Fixed Standards, Flexible Support

This point is directly connected to the last. In countries that expect their ninth or tenth graders to achieve at internationally benchmarked levels, we typically see that very few students are left behind and very few are pushed ahead by more than one grade a year. Virtually all but the special education students make a grade of progress for each year they are in school, against very demanding standards.

This requires very different supports for students than a system, like that of the United States, which is designed to operate by sorting students out along a long performance curve. In a system in which almost all students are expected to perform at high levels, the standard is fixed and the support varies to the extent needed to make sure that all students get to the finish line.

As we have already noted, this means that financial resources are allocated so that students who need more help are allocated more financial resources so they can get that help.



It also means that the students who are furthest behind get the best teachers, as is the case in Singapore. It is also the case in Singapore that the students who need help get more time, meaning time after school and on weekends and during the summer.

As we also saw above, in Finland and in many Asian nations, teachers are carefully trained to diagnose very quickly and accurately students who are beginning to fall behind and they are given the skills needed to figure out what those students need to get back on track quickly. In a sorting system, those skills are not very important, but in a system intended to get virtually all students up to a high standard and to keep them there, year after year, they are essential.

Low-Performing Schools

Sometimes it is not the student that is under-performing, but the school. This appears not to be a problem in Finland, where the variation in school performance is among the lowest in the world. As we have seen, Shanghai addresses this problem by requiring schools performing well to take responsibility for managing schools that are not

performing so well, by assigning high-performing staff members in high-performing schools to work in lower performing schools, by posting key staff members in low-performing schools to temporary assignments in high-performing schools to apprentice themselves to gain the skills they need, then sending them back to their home school and so on. Shanghai has also graded its schools by academic performance and the physical condition of its schools and shut down those in which both performance and physical condition did not justify continuation, sending the students and faculty to other schools as it built new schools to replace those in poor physical condition. Other Asian cities and nations have similar policies.

Design for Productivity

Management Paradigm

For many years, American policymakers have alternated between the search for quality and the quest for equity. What we are discovering is that other countries have figured out how to get both in greater measure than we. It would be natural for American educators to sigh and whisper that it would be wonderful to have both, but there is, apparently, no more money. Perhaps the most important discovery is that other countries have not only figured out how to get greater quality and far more equity, but they have figured out how to do that while spending substantially less than we do. They have not done it by doing a better job than we of managing the way we do. They have done it by adopting a very different way to organize the work of schooling.

The chief management guru of the early 20th century was Frederick Winslow Taylor. His counterpart for the latter half of the same century was Peter Drucker. Their messages were very different.

Taylor codified the methods of scientific management. Writing at the apogee of the mass production system, Taylor lived in a world in which goods and services formerly available only to the royalty and nobility were becoming increasingly available to Everyman, courtesy of very complex, very expensive machines that could turn out vast numbers of identical parts at remarkably low cost. Prior to the use of the mass production system, most finished products of any complexity were produced by craftsmen, one at a time, each object requiring great skill. But, in the mass production system, many fewer people—mainly the engineers who designed the machines and processes—needed high skills. Most other workers, from the people who minded the machines to those who assembled the parts into finished products to the clerks and the farm hands, required only basic literacy. Taylor declared that the way to run the system most efficiently was to observe many people doing these low level tasks, figure out who did them most efficiently and then make sure that everyone did it that way. Workers were just like the interchangeable parts they assembled. One was as good as another. Skill was not terribly important. Management just needed to make sure someone was doing the work and doing it efficiently.

The mass production method affected American industry more profoundly than that of any other major country. It was at its zenith when the current form of American

education was set in place. Though industry has long since moved on, the organization of work in American education has not.

Peter Drucker, in the 1970s opined that the age of mass production had reached its limit. The future, he said, belonged to firms and nations that embraced knowledge work and knowledge workers. By "knowledge work and knowledge workers," Drucker meant something very like "professional work and professional workers." Advanced industrial societies, Drucker said, would be able to maintain a high standard of living only if most of the workers were doing work that depended on them having a very high level of knowledge and the ability to apply that knowledge, case by case, to the challenges they faced every day. The challenges would be different, and so they would require a great deal of discretion as they figured out how best to respond to each challenge.

Taylor's methods would not work in such a situation. Workers would no longer be interchangeable. They would have to be managed in the same way professionals are managed and for the same reason. Rather than telling the workers just what to do and how to do it, managers would have to hire and train very high quality staff, set the goals, support the workers in every possible way and then get out of their way. The workers, who would themselves be the experts in the work, would have to figure out how best to meet the challenges they faced and would have to hold each other accountable for delivering top performance.

In the world of knowledge work, excellence would be rewarded. Blue-collar factory workers, Drucker said, expected an honest day's pay for an honest day's work. But knowledge workers, he said, expected an extraordinary day's pay for an extraordinary day's work like professionals in any field.

In varying degrees, all of the countries with high-performing education systems have been moving toward the management paradigm offered by Drucker. Few had embraced Taylor's system in its schools as avidly as the United States. But Taylor's paradigm is alive and well in American schools. It still influences our conception of teachers' work, the way we organize our schools, the way we talk about accountability, the way management in our schools relates to our unions, the way we respond to teacher shortages, the status of teachers colleges in our education system, and much, much more. Once the women and minorities who signed up for teaching when college-educated women and some minorities had a very narrow choice of careers retire, the United States is very unlikely to get the quality of teachers we need in the quantity we need them until we replace the Tayloristic paradigm of work organization with the model advocated by Drucker.

This is, of course, just what the top-performing education systems have been doing for years. The cases of Finland and Ontario are textbook examples of moves to forms of work organization in which teachers are treated much more like professionals and much less like blue-collar workers, cases in which management has been exercising progressively less control and providing progressively more support, and getting better and better results as a consequence.

Accountability and Autonomy

Accountability is one instance of the point just made. In Tayloristic management systems, the workers at each level are accountable to their supervisor. In many situations, as just pointed out, the worker is simply responsible for putting in an honest day's work for the requisite time on the clock. In others, the worker is paid by the number of units of product produced. In professional workplaces, however, while there is some element of accountability to one's supervisor, there is usually a major component of responsibility to one's professional colleagues for the quality and quantity of one's work. In professional workplaces, the workers are expected to put in whatever time it takes to get the work done. They feel a strong sense of responsibility to their colleagues to do their level best and they know that, at the end of the day, it is their colleagues, along with their supervisor, who will play a major role in determining their career prospects and very likely their compensation, both of which will depend on very nuanced judgments about their professional contribution to the work of the organization.

We can think of Tayloristic workplaces as emphasizing vertical accountability and professional workplaces as emphasizing lateral accountability. In Tayloristic workplaces, it is always very clear who the workers are and who management is. In professional workplaces, it is often the case that the professionals are organized as a partnership, and the workers are also the managers as well as the owners. Even when this is not the case, there is typically a strong element of lateral accountability in professional workplaces and it is usually also the case in professional workplaces that the workers are also managers, though they may not also be owners.

These differences in accountability between Tayloristic management systems and professional systems are a function of the nature of the work. If the work can be done by semi-skilled people who are essentially interchangeable and whose work is most efficiently managed by supervisors who are in a position to direct the work in detail by virtue of their superior knowledge, then a top down system of accountability will probably work best. But if the work is of the kind that Drucker was interested in, then the people in the best position to make the judgments about the way the service will be delivered will be the people actually doing the work, and they will have to have a wide range of discretion in determining how it will be done. The incentives that work in a Tayloristic workplace will not work in a professional workplace. Professionals, as Drucker pointed out, are much more motivated by the need to excel in the eyes of their professional colleagues and to meet professional norms. They will do whatever it takes, knowing that, if they don't, they could lose not only their job, but also the respect of colleagues whose respect they greatly value.

The other side of increased lateral accountability is increased professional autonomy. When there is one best way to get the work done, the job of management is to make sure it gets done that way, but when the best way to get the work done is a function of the particular unique situation one faces, then the professional must be free to make the decision as to how the service will be delivered to the client. One way to frame this is to

say that management has little choice in that situation but to trust the professional to know what to do and to do it.

But schools are small societies, not collectives in which each professional is an individual entrepreneur. Some teachers are better at one aspect of the overall work than another, just as some attorneys are better at bringing in new clients and others are better at research and writing and others are better at litigating. The law firm works best when these different skills and abilities are welded together in one team. So it is with a school. In such a situation, it is the senior members of the workforce who are in the best position to judge the contributions of each member of the team. Each has plenty of professional autonomy, but each is responsible to the other members of the team for the quality and timeliness of their work.

There is a general trend among the countries with the most successful education systems away from Tayloristic models and toward the kinds of accountability systems associated with professional work. The Japanese emphasis on earning the respect of the group of which one is a part puts great pressure on Japanese teachers to be accountable to the rest of the faculty for the effort they put into their work and the quality with which they do it. In recent years, the Ministry has, somewhat cautiously, begun to provide progressively less explicit direction to the schools and to provide greater degrees of freedom to school faculties with respect to how the Japanese curriculum will be implemented and on other matters. We can see similar trends in Singapore and China.

The Finnish reforms in the 1970s resulted in a much-admired and rather detailed specification of the Finnish curriculum. But, in the period since then, there has been a steady reduction in the detail with which the curriculum has been specified and the Ministry has abolished the Finnish inspectorate. All this has happened in a country in which there are no national examinations of all the students, so that neither schools nor teachers can be held accountable for their performance on the basis of data from such examinations. All of these policy positions are a measure of the high degree of trust that the Finns have in their teachers, but the high performance of Finnish students is a testament to the degree to which Finnish professionals hold each other accountable for the quality of their work and the effort they put into it.

The Canadian province of Ontario is another case, much like Finland, in which the current administration has abandoned the policies of its predecessor in favor of a policy of providing great discretion to teachers and trusting them to do the right thing, and getting great improvement in student performance in return.

Incentives

The way incentives are structured can make a big difference in the relative productivity of systems. Perhaps the best example is the effect on student motivation of the use of external examination systems as gateways by the best-performing nations. In countries with external examination systems used as gateways, as we noted, students have strong incentives to take tough courses and work hard in school. In the United States, unless a student is headed for a selective college, he or she quickly realizes that, even if the

objective is to get into an open-admissions college, it makes no difference whether the student gets good grades or a D minus, the result is the same: entrance to a non-selective college. The effect is to send a message to our students that high school is a place to hang out with one's friends. As long as you show up, you will do as well as you would if you take school seriously. What they do not know, of course, is that, if they have not done well enough to succeed in their initial credit-bearing college courses, they will have to take remedial courses for which they will receive no credit, while piling up debt. By the time they learn that, it is too late.

American policymakers assume that all school faculty have positive incentives to adopt research findings that show X works better than Y. But that is not true if they think that adopting X may arouse the anger of some vocal group in the community. Administrators are almost certain to get into deep trouble if they take high cost contracts away from local contractors in order to give them to lower cost national contractors, even though doing so would save a lot of money that could be used for instruction. Actually, faculty have stronger incentives to avoid trouble than they do to do what works for students. School people have no incentive to meet the needs of minority and low-income students if their performance improves and the money is taken away. If school administrators find a way to deliver the same services for less money, their reward is to have their budget reduced. Education school deans report that, if they propose to raise standards for admission in their schools, the arts and sciences faculty may veto that move because it might mean fewer students in their departments. Some minority students in inner city schools who decide to work hard in school are turned into pariahs for "acting White." Some teachers who do whatever it takes for their students are ostracized by their colleagues for violating the union contract. Teachers who teach complex skills to their students that are not measured on the standardized test they must give are sometimes penalized because they are not sticking to the schedule for teaching much lower basic skills. These are all examples of perverse incentives, that is, positive incentives for lowering, not raising, achievement. Our education system is rife with such perverse incentives.

High-performing education systems typically have far fewer perverse incentives than the American system. We have already pointed out that all students, not just those going to selective colleges, have strong incentives to take tough courses and study hard in the top-performing countries. Teachers in Japan have strong incentives to work hard and perform at high levels because of the value that all Japanese work groups place on that behavior. The Singaporeans provide substantial bonuses to teachers to do outstanding work. Teachers colleges in the best-performing countries are not expected to be "cash cows" for the arts and sciences schools in those countries. And so on.

If one does not like the performance of the education system, it is easy to blame the actors. But the chances are that you would behave just the way they are behaving if you were experiencing the same incentives. If you want better performance from the system, one of the first places to look for opportunities is the structure of incentives in that system. If you find a lot of perverse incentives—incentives to produce the behavior you do not want—then change the incentives. Our best competitors have done just that.

School-to-Work Transition

Investing more in education is sort of a bet, a bet that giving students a better education will result in certain outcomes. Among those outcomes is that they are more likely to be able to support themselves and their families and enjoy a good standard of living. But there is no direct connection between being well-educated and earning a good living. Students need to make an effective transition from school to work and that process is more complicated than it might at first appear.

Among other things, it involves turning academic skills into the kind of skills that are needed to do particular jobs, which always involves more learning, a part of which usually takes place on the job, under the supervision of an experienced hand. It involves an opportunity to get that experience, which usually requires access to an informal network of people who have jobs, internships or apprenticeships to offer. And it involves the acquisition of many skills and kinds of knowledge that are not included in the usual school syllabus.

Some countries have effective systems to effectuate such transitions and many do not. The United States is among the latter. Many graduates in the United States have few, if any, family connections to people who can and will offer them the first rung on the ladder, the chance to acquire the initial experience needed. Many lack the specific skills, attitudes and dispositions needed to succeed in those jobs. The result is very high youth unemployment rates, a high rate of youth delinquency and crime, and ruined lives.

Finland has multiple pathways that are highly developed and successful at delivering occupational skills at the upper secondary level, as does Singapore. Japan reaches much the same goal through its system of having designated high schools that supply high prestige employers with high quality candidates, who are then provided very high quality on-the-job training in the quality circles operated by those firms. These systems are very different from one another, but each is a vital component of that country's system for providing a rewarding future for all its children and a capable workforce to drive its economy. The point here is that a country may have a high quality pre-college education system and still have a low-quality workforce if it fails to create a sound school-to-work transition system.

Single Capable Center

Every high-performing country the National Center on Education and the Economy has studied has a unit of government that is clearly in charge of elementary and secondary education. In Canada, those units of government are not at the national level (the national government has even less responsibility for the schools than the federal government in the United States) but at the provincial level. In Finland, Singapore and Japan, it is the national Ministry of Education that is in charge. In China, Shanghai has unusual independence from the national Ministry of Education.

In many of these countries, educators view a position in the ministry as the capstone of a distinguished career. The ministry sees itself, and is seen by others, as having great

legitimacy as the keeper of the whole system, the agency responsible for defining the future course of education and for leading the national discussion as to the best shape for that system. It is often the case that these ministries do not have to issue many regulations because their informal guidance is so respected.

In such countries, the ministry has an obligation to concern itself with the design of the system as a whole, with the structure of incentives that design provides to everyone affected by it, with the coherence of that design and with the ability of that design to address the problems the country faces.

No unit of government in the United States occupies such a position. No one expects or wants the US Department of Education to play that role for the United States. Certainly, no city school district plays the role just described. But it is also true that no state department of education has a role comparable to that of a typical national ministry of education.

That is not because our state departments of education lack the constitutional authority to play that role. Most state departments of education are required by their state constitutions to provide a 'thorough and efficient education' to their citizenry. But two centuries of practice have vested a great deal of authority in local boards of education, to a degree that has no parallel in most other countries, and that authority was essentially delegated from the state a long time ago.

The result is that no level of government in the United States thinks of itself or is thought of by others as the place where the buck stops, the place where responsibility ultimately resides for the effectiveness and efficiency of the system as a whole. And the result of that is that education reform in the United States takes a different form than it typically does in the countries with the most effective education systems. When compared with other countries, the United States appears to see education reform as a process of adding programs to the corpus of programs already in place. We endlessly initiate new programs in the announced hope that they will somehow prevail, but the reality is that they gain favor with early adopters and rarely go much further. Where other countries carefully consider new policies and work hard to integrate them with existing ones in ways that will increase rather than decrease system coherence, the United States simply adds another program and hopes for the best. Which leads directly to the next point.

On Systems, Coherence, Alignment and Tradeoffs

It is at this point that the author will peep out from behind the screen of the anonymous voice and speak in the first person. After 22 years of research on the factors that account for the success of the countries with the best education record, I find myself convinced that seven things account for the lion's share of the difference: 1) aggressive international benchmarking, 2) the quality of the teaching force, 3) the use of aligned instructional systems and external examinations that measure complex thinking skills, 4) the decision to get all students to those standards, 5) the use of professional systems of work organization instead of blue-collar models, 6) funding systems that put the most funds behind the students who are hardest to educate, and 7) coherence of the design of the

overall education system itself, in all of its particulars. If I were forced to reduce the list even further, I would choose the second and last of these (though equitable funding is a close runner-up).

Coherence of system design is that important. Why this is so is not immediately obvious. Our education research tradition has taught us to think in terms of the effectiveness of individual initiatives. We use statistical techniques to create a virtual environment in which we can simulate the effect of the intervention of interest on the outcomes of interest, everything else being equal. Then we wonder why the effects of even the most powerful interventions are almost always trivial.

The reality is that the outcomes we care about in education are the result of myriad variables, all jostling with each other in a great vat, interacting in ways we can not possibly visualize or simulate in our computers, to produce the outcomes we see. Each program we evaluate with our sophisticated research techniques can actually be considered in real schools and school systems as one among many variables affecting the outcomes we care about. If no one thinks of themselves as responsible for the design of the overall system of which those variables are a part, then we should not be surprised that any single initiative or program, no matter how well conceived and executed, has a relatively small effect on student achievement. Because so many things affect the outcome, in ways that no policymaker has thought very much about, it is to be expected that altering one variable cannot affect the outcome very much at all, one way or the other. The one thing that could have a very large effect—the design of the system itself—is no one's responsibility.

Visiting the average school is a bit like an archeological exercise, consisting of unearthing layer upon layer of initiatives carefully deposited in the school over the decades of its existence: a text that the social studies text selection committee liked ten years ago when it was all the rage, an instructional method that Jack and Judy brought back from their professional development program during the last administration, that technique that the central office was onto six years ago and caught the fancy of our then-principal, who of course moved on last year and was replaced by a principal with a very different agenda. But none of it ever really goes away. Legislators add law after law, the courts make their decisions, the state department issues regulation after regulation—all of it is added on until it looks like the folded sedimentary rock in the road cut on the interstate going out of town.

It is little wonder that our systems are full of negative and perverse incentives. No one ever thought about how all of these layers of law, regulation, court decisions, textbook choices, professional development programs and much, much more fit together and so it is little wonder that they do not. As we pointed out above, the texts do not align to the curriculum, which are not aligned to the assessments, which are not aligned to what teachers are taught in teachers colleges, which is unrelated to the curriculum frameworks, which do not exist.

Americans can only imagine what might happen if we had an education system in which the parts and pieces of the system were constructed to fit together in a sensible way, so that they reinforced each other rather than spent their lives fighting with each other. This is the end result of living in a country that was founded by people who deeply distrusted government and believed that education was one arena in which local decisions would be best, because local people knew best what their children would need to be successful.

The reality is that local control is mostly honored in the breach. Textbook manufacturers control the curriculum actually taught, to the extent that anyone does. Districts must choose among national tests made by national testing companies. The curricula of schools of education are more influenced by the curricula of other schools of education around the country than by the state in which they are chartered. Local control is a chimera. But no one else is in control either.

Our forefathers and foremothers never imagined a world in which the sons and daughters of local citizens would be competing for jobs directly with the sons and daughters of people who lived on different continents in a very complex global economy that would require highly complex education systems designed and overseen by people with rare expertise. But that is the situation we now face and our educational institutions are not well equipped to cope with it.

To talk with the people who run the Singapore education system is to hear a tale in which the designers worked as an engineer would work to build an ever more effective system, step by step. That is actually just what they did, rising from third world status fifty years ago to front rank status today. Wave on wave of visitors have descended on Finland to find out what key policy initiative vaulted them to world class status while no one was looking. But the visitors find out that there was no single policy initiative the Finns took to get where they are. Like the Singaporeans, the Finns, it seems, worked in a logical way, while governments came and went, in small increments over the same fifty years to take an education system designed to support a small rural economy to world leadership in just five decades. At each stage, these countries had education systems that were genuine systems.

It is only when one considers the education system as one coherent whole that it becomes possible to analyze and deal with the tradeoffs that are inherent in any system.

Consider Japan, for example, where, as we have seen, the overall ratio of students to teachers is much the same as in the United States, but the classes are considerably larger, leaving much more time for teachers to plan and develop more effective lessons and to work with individual students and small groups of students.

Consider Finland, where the government has provided its teachers with greater autonomy with respect to the curriculum and accountability as the quality of its teachers have improved. Reducing the detail with which the curriculum is specified, virtually eliminating test-based accountability and closing down the inspectorate, which is what the Finns have done, would make no sense at all if the Finns had doubted the quality of

their teachers, but all became necessary when they had managed to produce one of the highest quality teaching staffs in the world. It is essential for a high-performing country to trust its teachers, but it had better have teachers it can trust.

The most important tradeoffs undoubtedly lie in the area of system effects having to do with investments in quality.

The American mass production system was primarily concerned with driving cost down as low as possible. Quality was secondary. American production lines would produce a lot of parts and finished products that needed to be thrown out or remanufactured. But, in the latter half of the 20th century, the Japanese, borrowing American ideas that did not get a hearing in the United States, started to reengineer their manufacturing systems to assure that quality was built in at every stage of the process, with the result that the finished product met very high quality criteria with very little wastage produced along the way. They actually showed that it is less expensive to build quality in at the beginning than to compensate for the lack of quality at the end of the production line.

Part of the price paid by the American education system for being built on the mass production model is that we tolerate an exceptionally high rate of wastage. Only in our case, what is being discarded is young people. We see this in the very high percentages of young people who are not fluent readers by the time they leave elementary school, the very high rates at which students drop out of high school, the appalling rates at which those who enroll in college need remedial work when they get there and the equally appalling rate at which they drop out and never receive a degree.

That does not happen in the countries with the best-performing education systems. These countries have learned how to build quality in beginning before birth and extending throughout the entire education process. One illuminating example will suffice. The United States, as we explained above, is now bottom fishing for its teachers, sending them to low status training institutions, preparing them poorly for teaching, not supporting them in their initial years while they are learning the ropes and compensating them poorly. It should not surprise us that a significant number of teachers do not do a good job, nor should it surprise us that many want out. Close to a third of those who trained as teachers are gone within three years and close to half are gone in less than five years. These rates are significantly higher than for other occupations.

Imagine what would happen if they stayed for ten years, on average, instead of three to five years. We would need fewer than half as many slots in our teachers colleges. We could afford to upgrade their training substantially and still have money left over, which we could use to provide them with better support when they get their first job and there might even be money left over to raise their pay. We might be able to get a world class teaching force for the same money we are paying now, in the same way that our automobile companies found out that they could produce much higher quality cars for the same money it cost to produce low quality cars.

Consider another take on the same theme. As noted above, most of the top-performing countries are getting their students through the common curriculum by the end of the lower secondary school, or about the age of 16. We shoot for the doing the same thing by the end of upper secondary education. Suppose we set our system up to match their achievement. We could save the cost of the junior and senior year of high school. Of course, we would not really save it, because we would need the extra money to make the improvements needed to get all our students to the goal line by the time they are 16. But the reality is that 30 percent of our students drop out, and a substantial fraction of the rest leave high school with no more than an eighth or ninth grade level of literacy. Our competitors have dropout rates in the neighborhood of 10 percent or less and they leave with average literacy rates far higher than ours. So we could get much better results than we are getting now for the same money by taking the money we are wasting on the last two years of high school and spending it wisely in the earlier years, as our competitors do.

The reason I believe that high quality staff, equitable funding and coherent systems are the key to highly successful education systems is that these points lead to all the others. Any country that recruits its teachers from the higher ranges of the applied ability distribution will quickly find that—in order to keep them—it has to train them in high quality, high status universities, support them well once hired and offer them decent pay and professional work environments, and—not least—trust them to do the right thing. Any country that really strives for coherence and which seriously researches the best practices of the leading countries will in time be forced to adopt high quality curriculum-based examinations and use them to define a few important gateways, to develop strong curriculum frameworks, and to fund their schools equitably and make sensible trade-offs as they make decisions about how their money will be spent. Any country that moves toward a system of truly equitable school finance has made the crucial decision to get all of its students to high standards. These key practices, if informed by serious international benchmarking will, in time, lead to all the others.

The Dog That Did Not Bark

In one of Arthur Conan Doyle's best-known Sherlock Holmes stories the clue is a dog that did not bark. In this case, the dog that did not bark is the dominant element of the American education reform agenda. It turns out that neither the researchers whose work is reported on in this paper nor the analysts of the OECD PISA data have found any evidence that any country that leads the world's education performance league tables has gotten there by implementing any of the major agenda items that dominate the education reform agenda in the United States.

We include in this list the use of market mechanisms such as charter schools and vouchers, the identification and support of education entrepreneurs to disrupt the system, and the use of student performance data on standardized tests to identify teachers and principals who are then rewarded on that basis for the value they add to a student's education or who are punished because they fail to do so.

This is not to say that none of these initiatives will lead to significantly improved performance at scale. It is only to say that none of the countries that have the best records of performance have employed these strategies to get there.

It is important here to make it clear that many countries are interested in current efforts in the United States to identify through research what makes for good teaching and for a good teacher. They understand that such information, if we can get it, would be very useful in creating criteria for admission to high-quality teacher education programs, for designing those programs, for producing better criteria for licensure, for creating better professional development programs and for evaluating teachers. But they worry that using standardized test data as a major basis of evaluating and rewarding teachers will create perverse incentives of many kinds and they also worry both that there is much in student performance that is important that standardized tests are unlikely to capture and that great student performance is the result of the work of many adults working in collaboration rather than individual teachers working alone.

An Agenda for American Education

What follows is a new agenda for recasting the structure of the preceding section, derived from the experience of the countries that have consistently outperformed the United States. It was constructed simply by taking the subsection headings and reframing the language of the preceding sections in the form of an action agenda. To be clear, this is not an agenda for the United States; it is an agenda for individual states:

- **Benchmark the Education Systems of the Top-Performing Countries**
 - Make sure you know what the leaders are trying to achieve, the extent to which they achieve it and how they do on common measures
 - Compare your state to the best performers, with particular attention to countries that share your goals
 - Conduct careful research on the policies and practices of the best-performing nations to understand how they get the results they get
 - Benchmark often, because the best never stand still
- **Design for Quality**
 - Get your goals clear, and get public and professional consensus on them
 - Create world-class instructional systems and gateways
 - Define a limited number of gateways — not more than the end of basic education, end of lower secondary and end of upper secondary (matched up to college entrance and work-ready requirements)
 - Create standards for each gateway, making sure they are properly nested and are world class

- Create logically ordered curriculum frameworks (topics for each year for each subject) for the basic education sequence
 - Create curriculum (broad guidelines, not lesson plans) for each school level leading up to the gateway exams (the level of detail at which this is done should be inversely related to the quality of your teachers)
 - Create exams for each gateway, based on standards and curricula
-
- Train teachers to teach those curricula well to students from many different backgrounds
- Develop a world-class teaching force
 - Raise standards for entry into teacher education to internationally benchmarked levels, including standards for general intelligence, level of mastery of subject matter content and ability to relate to young people, with rigorous selection processes
 - Move teacher education out of second and third tier institutions and into the major research universities
 - Insist that teachers of all subjects at all levels have a depth and breadth of mastery of the subjects they will teach comparable at the bachelors degree level to that of the people who will go on to graduate education in those fields
 - Make sure that prospective teachers have excellent skills in diagnosing student problems and prescribing appropriate solutions
 - Design the teacher preparation program on a clinical model, with plenty of clinical experience under the constant supervision of master teachers in real settings
 - Raise the criteria for teacher licensure to internationally benchmarked levels and never, under any circumstances, waive the licensure standards in the face of a teacher shortage
 - Make sure compensation for beginning teachers is and remains comparable to compensation for the other non-feminized professions; add the amounts necessary to attract capable teachers to hardship locations, and specialties in shortage; tie amounts to steps on the career ladders (see below)
 - Provide for an induction period for new teachers of at least a year in which they are supervised by master teachers who are released from full time teaching for this purpose
 - Construct multiple career pathways for teachers one of which is into school administration, at least one of which is in teaching and all of which provide for merit-based advancement with increasing responsibility and compensation

- Set up a system for identifying teachers who have been in service for a few years who have the attributes likely to enable them to be strong candidates for one of the career pathways; groom them for advancement by offering them free advanced training tied to the steps on the career ladder; provide mentoring and other forms of support and continue that support as long as they continue to be promising candidates for advancement.
- Explore the development of approaches to instruction that would enable the state to get world-class results with larger class sizes. Class size is important because it is the fundamental driver of teacher cost and teacher cost is the fundamental driver of the cost of the entire system. Japan has shown how it is possible to increase class size and increase student performance at the same time. Perhaps that method would work in the United States, perhaps not. It is important to find out and, if it does not work or work as well, to make as much progress on this front as possible.

- **Design for Equity**

- Move toward full state adoption of responsibility for school finance and toward implementation of a weighted pupil finance system, which would calculate the amount due each school entirely on the basis of a uniform state formula. Let parents and students choose among public schools, with the funding following the student. The formula would provide funding to any public school chosen by the parents and the student, with the same base funding behind all students in the state, but additional amounts going to students based on the cost of bringing that student up to the high state academic standards. Among the students bringing more money to the school would be those from low-income families, students from families that do not speak English at home and those with some form of disability.
- Develop a system in which all schools, from kindergarten through the end of lower secondary school, are truly comprehensive, open to all children of all races, ethnicity, gender and socio-economic status and are untracked, and committed to bringing all students up to the same high standards, irrespective of background.
- Make sure that schools have the same high expectations for all students and that they provide the additional supports required by students who need them to achieve those standards (which is why a weighted student formula for school funding is necessary)
- Identify schools that are not succeeding in bringing all their students to high standards and close those schools and distribute the students to high-performing schools, send key staff from better-performing schools to take leadership positions in the low-performing schools, and send key staff from low-performing schools for training in the high-performing schools or have the managements of high-performing schools also take responsibility for managing the low-performing schools.

- **Design for Productivity**

- Adopt as a conceptual framework for the reform program the goal of reframing teaching from a feminized occupation performed in a Taylorized work organization to professional work (or knowledge work, as Peter Drucker would have it) performed in a form of work organization appropriate for professionals
- Look for opportunities to build quality into the education system from the beginning rather than cope with the high rate of wastage in the current system
- Examine the total state budget for opportunities to make better tradeoffs between major budget elements in favor of higher productivity
- Do what is necessary to redesign the state department of education so that it has the capacity and status needed to drive the state education system to excellence
- Examine the state's school-to-work transition system to see if it is truly world class in the way that it enables all young people who want it to get access to high quality work experience and on-the-job training, access to networks of people who are offering good jobs and access to further schooling designed to provide high quality education and training leading to industry-recognized occupational certification.

- **Make sure your systems are coherent and aligned**

But that's impossible! Realistically, how can we get started?

Sure, you say. All this sounds sensible and you have explained that it is all being done somewhere by somebody, but it simply cannot be done here, in these United States, or at least in my state, in the foreseeable future. Too many vested interests, too deep a commitment to local control, too many teachers colleges to be shut down, too many objections from unions, too few master teachers available, just too much!

It has taken from 30 to 100 years to build the national and provincial education systems on which these recommendations are based. None were built in one or two decades. If the United States is to catch up, it will have to get started soon and will have to work very hard at it for a long time. But what to do while waiting for the long-term payoff?

We have not mentioned Canada much until now, because this is where it fits. The government of Ontario did not predicate their reform program on replacing its current teacher workforce with a new workforce. They did not think they needed to. They asked themselves how they could get much better results from the workforce already in place. The answer they came up with was to make peace with the teachers unions that had been demonized by the previous administration and with the teachers that had been so badly

demoralized and they invited them to join them in thinking through a reform program that would improve student performance. They insisted on high standards but they listened hard to what the teachers had to say about the support they needed to raise student achievement to those standards. They decided that the highest leverage strategy available to them was to build the capacity and professional skill and commitment of their in-place teaching force. They focused on what it would take to build capacity at every level of the system to deliver, and wherever possible, supplied it. They made a point of trusting teachers and the teachers returned their trust.

Earlier, as we have also related, they redesigned their school finance system to create one far more equitable than the one they had had. It is impossible to overstate the importance of this policy change. On that foundation, they built an education system, province by province, that put the nation as a whole comfortably among the top ten performers in the world.

The measures just described did not result in equal improvements at all student ability levels. There was broad and substantial improvement for the students in the bottom half of the achievement distribution, but much less among those who had been doing better before these measures were introduced. There was considerable improvement on measures of basic skills, but nowhere near as much on measures of higher order skills. Which is exactly what one would expect of such an approach. It is not surprising that, with the same teachers in place who had been in place before these initiatives, and with a strong effort to build capacity in the teaching force where the teaching force felt it was most in need of additional capacity, one would see the most improvement among the students who had been doing least well.

One way of looking at what the Ontario government did was that, by building the capacity of the current teaching force, they took the distribution of student performance and moved the left tail of the performance curve toward the middle of the curve, while the middle and right hand parts of the curve did not change much. One can think of their next challenge as moving the entire curve to the right, so that the performance of all students improves substantially, and the performance of the students who perform least well is not far from the best-performing students, who would then be performing at world class levels. That is precisely how we defined world class performance at the beginning of this paper. To get that, we would argue, Canada would have to adopt the other features of the agenda of their top-performing peers.

And that is exactly what we think makes sense in the United States. Start with the Canadian agenda, while also, at the same time, begin to work on those parts of the larger agenda that seem possible at the outset. The strategies chosen would be different for different states, depending on what is politically possible, what the state's strong points are and the nature of its weak points. But working over time in this way strikes us as plausible in the real world.

Bear in mind, we are not suggesting that it is possible to short cut the steps the top performers have taken on the way to the top of the league tables. Canada, like many of

the other top performers, has moved the preparation of its teachers into the universities. In order to teach in Ontario schools, high school graduates must complete a degree program in the subject they wish to teach and another degree program lasting at least a year in professional education. This includes elementary school teachers, who must specialize in one or two subjects in the elementary curriculum, such as English, history, science or mathematics. Secondary school teachers must have academic credentials in at least two subjects, such as English and history, or music and mathematics. Candidates who think they might want to be a subject specialist must take an honors degree. High school students must have 3.2 to 3.3 grade point averages on a scale of four to get into the institutions offering the first of these two degrees. There are fewer universities per capita than in the United States and the universities in which teachers are trained have a higher status than their opposite numbers in the United States. Teachers in Canada are better paid than American teachers.

It might be fair to say, then, that the Canadian benchmark before embarking on the current round of reforms was above where the United States is now, but within reach. An American state could reasonably set an agenda for reaching toward the Canadian starting line, then their current state and then the more distant configuration of public policy for education that has been adopted by the very best performers in the world. That is a very ambitious agenda, but it is doable, by stages.

What the Federal Government Can Do

No one wants a national education system in the United States. Even if one wanted to mandate that a state adopt an agenda of the sort described above, it would not work. The kinds of systems we described would not be faithfully implemented in a state that was opposed to them, no matter what compliance mechanisms were used. Nor is it very likely that all states would want to embark on such an agenda. That logic suggests a federal government interested in the adoption of such an agenda would be well advised to provide assistance to states that would really like to implement such an agenda, but which, in the current environment, lack the resources needed to do so.

The agenda we have laid out here is consistent at many points with the markers that the Congress and the Obama Administration have already put down. This paper began by noting that Secretary Duncan has reversed half a century of history by actively calling the attention of this country to the achievements of the countries that are outpacing us in education and doing something to learn how they do it. The Race to the Top program was designed and passed in a form that encourages the kind of comprehensive and coherent planning advocated here, rather than the digging of potholes encouraged by categorical programs. Through the Common Core State Standards work, a major step toward the implementation of the kind of internationally benchmarked standards embraced by all high-performing countries was initiated by the states, and has received the enthusiastic support of the Administration. And the Administration initiative to use Race to the Top funds to support the development of tests matched to the standards should move the United States much closer to the kinds of powerful, cohesive instructional systems the top-performing countries have. The President's call for making

all high school students college and career ready and for setting a goal of once again leading the world in college completion is a big step toward developing the kind of consensus on education goals that characterizes the countries with the best education performance. And the Administration has proposed a number of initiatives on teacher quality in the United States that are consistent with the strategies other countries have taken to assure themselves a strong supply of high quality teachers in the years to come.

~~So the stage is set. The time has come to build on these beginnings and to embrace aggressively a comprehensive agenda that is squarely based on the principles that lie behind the success of those countries that have been leading the world's education league tables.~~

This paper is being written on the eve of reauthorization hearings for the Elementary and Secondary Education Act. We suggest that a title of that act be written that would create a competition among states for funds that would be used to implement the agenda described in this chapter. We would make sure that there was considerable latitude for the states in the way they approached their design for implementation. It might be appropriate for the federal government to conduct activities intended to broadly familiarize the states with the strategies being employed by the countries with the most successful education systems before the competition takes place. People familiar in detail with those strategies, including representatives of the countries at the top of the league tables, people who have researched those countries, as well as people familiar with each states' current situation, might be involved as reviewers of the state proposals. After the first round of such grants is made, the government might wish to sponsor additional rounds.

We would be leery of mandating specific design features in the announcement of such a program, much less implementation schedules and deadlines. States should be free to build on their existing strengths and to minimize their weaknesses as they build their strategies. Their strategies need to reflect their politics and their history. The review process ought to be less a compliance check than an assessment of their determination and their capacity to take full advantage of the path blazed by the countries with the most successful education systems. Let the states convince the readers that they understand what has happened in these countries and are prepared to do what is necessary to adapt and profit from that experience in their own unique ways.

What the States Can Do

But the real action would be, of course, in the states. Whether or not the federal government chooses to take an active role, the states have all the authority they need to move in the direction outlined here. This is, needless to say, a very ambitious agenda. It is inconceivable that it could be successfully implemented without capable and determined leadership to produce a wide consensus for the main outline of the work. In almost every case described in this paper, there was an individual or a political party that provided unusual continuity of leadership for this agenda over a long period of time. That is not easy to achieve in the United States, but not impossible, either.

The claim that this agenda has on our attention is simply that it has worked. It has worked in countries as different as Singapore and Finland, Japan and Canada. It is not a Republican agenda or a Democratic agenda. It is neither conservative nor liberal. While it requires major changes in the way we do things in the United States, it demands changes more or less equally of all parties. The changes it calls for are as dramatic as the changes made in government in the Progressive Era, but let the record show that the United States made those changes. It can make these, too, if it chooses to do so.

Bibliography

Allegretto, S. and Mishel, L., *The Teaching Penalty: Teacher Pay Losing Ground*, Economic Policy Institute, Washington, DC, 2008.

Asia Society, *Canada: School Improvement without Rancor and Rankings*, New York, NY, 2010.

Auguste, B., et al., *Closing the talent gap: Attracting and retaining top-third graduates to careers in teaching*, McKinsey & Company, September 2010.

Costrell, R., and Podgersky, M., *Teacher Retirement Benefits*, EducationNext, Cambridge, MA, volume 9, number 2, Spring 2009.

Education Trust, *CORE PROBLEMS: Out-of-Field Teaching Persists in Academic Courses and High-Poverty Schools*, Washington, DC, November 2008.

Economic Policy Institute, *How Do Teachers Wages Compare? Methodological Challenges and Answers*, Washington, DC, 2004.

Franke, R., et al., *Findings from the 2009 Administration of the College Senior Survey: National Aggregates*, Higher Education Research Institute Graduate School of Education & Information Studies University of California, Los Angeles, CA, February 2010.

Gitomer, D., *Teacher Quality in a Changing Policy Landscape: Improvements in the Teacher Pool*, Educational Testing Service, Princeton, NJ, 2007.

Heilig, J. and Jez, S.J., *Teach for America: A Review of the Evidence*, The Great Lakes Center for Education Research and Practice, East Lansing, Michigan, 2010.

Higher Education Research Institute, *Findings from the 2009 Administration of the College Senior Survey (CSS): National Aggregates*, Graduate School of Education and Information Studies, UCLA, Los Angeles, CA, February, 2010.

Hurley, E., *Teacher Pay 1940—2000: Losing Ground, Losing Status*, National Education Association, Washington, DC, 2006.

Lykins, C. and Heyneman, S., *The Federal Role in Education: Lessons from Australia, Germany, and Canada*, Center on Education Policy, Washington, DC, November 2008.

Markow, D. and Cooper, M., *The MetLife Survey of the American Teacher: Past, Present and Future*, MetLife, Inc., New York, NY, October 2008.

MEXT: Japan Ministry of Education, Culture, Sports, Science and Technology, *Education Reform for the 21st Century, National Council of Education Reform and the Process of its Reform*, Tokyo, Japan, 2010.

Ministry of Education, *Guidelines for Curriculum Reform in Basic Education*, Beijing, China, 2001.

Ministry of Education, *Towards Thinking Schools*, Singapore, 1997.

Ministry of Trade and Industry, *The Singapore Economy: New Directions*, Singapore, 1986.

National Association of Colleges and Employers, *Salary Survey*, Bethlehem, PA, 2010.

National Center for Education Statistics, *Characteristics of Schools, Districts, Teachers, Principals, and School Libraries in the United States: 2003-04 Schools and Staffing Survey*, Washington, DC.

National Science Foundation, *Science and Engineering Indicators: 2010*, NSB 10-01, Arlington, VA, 2010.

Norrie, K. and Lin, S., *Postsecondary Education Attainment and Participation in Ontario*, Higher Education Quality Council of Ontario, Toronto, Ontario, November 2009.

Ontario College of Teachers, *Thinking About Teaching*, Toronto, Ontario, 2011.

Ontario Focused Intervention Partnership 2007-2008, PowerPoint Presentation, The Literacy and Numeracy Secretariat, Ontario Ministry of Education, Ontario, Canada, downloaded March 2007.

Orland, M., *School Turnaround Policies and Practices in Australia, Canada, England, and New Zealand: Overview and Implications*, San Francisco, CA, WestEd, March 2011.

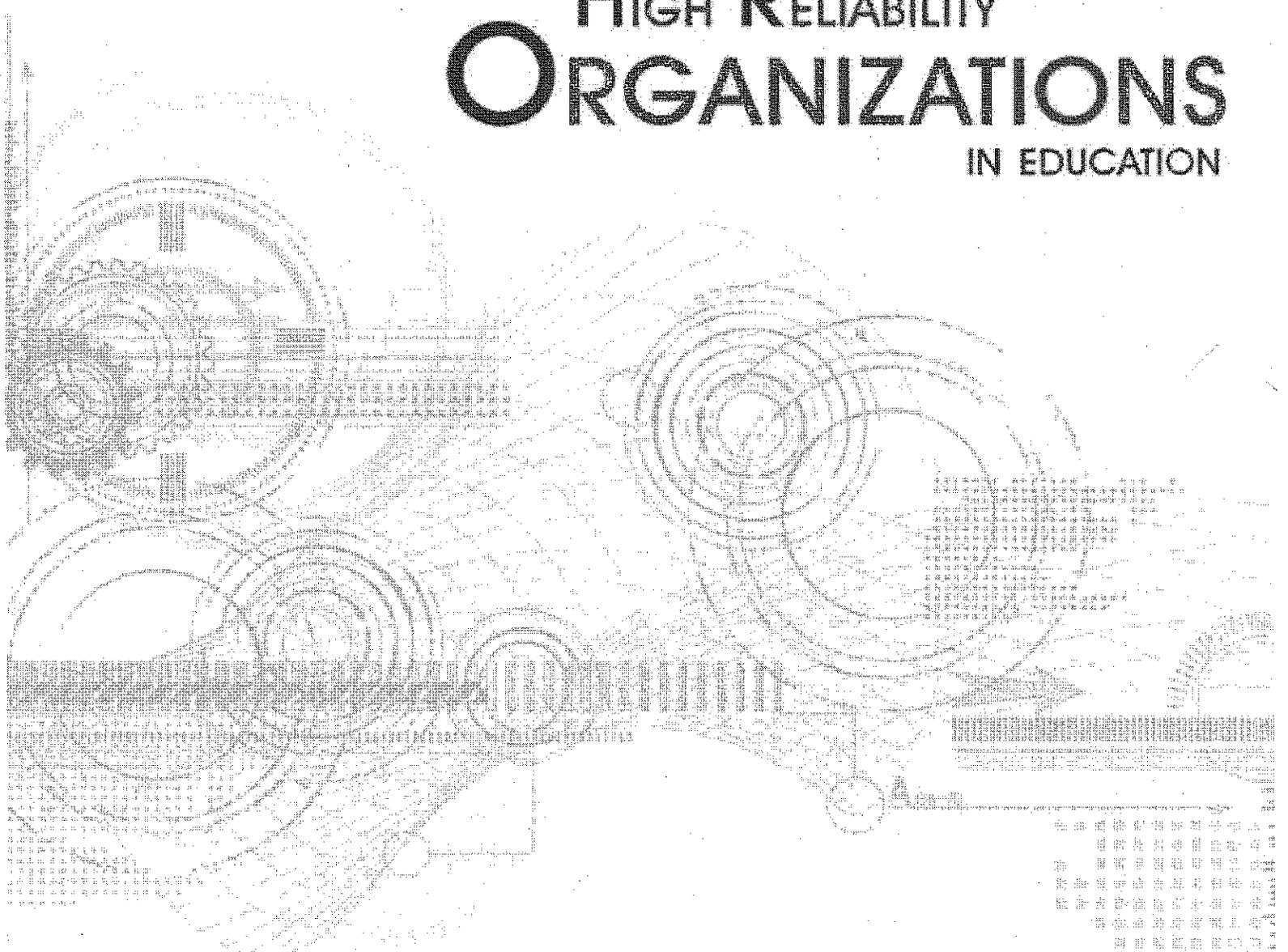
Sclafani, S. and Lim, E., *Rethinking Human Capital in Education: Singapore As A Model for Teacher Development*, The Aspen Institute, Washington, DC, 2008.

The College Board, *2008 College-Bound Seniors, Total Group Profile Report*, New York, NY, 2008.

Noteworthy

PERSPECTIVES

HIGH RELIABILITY ORGANIZATIONS IN EDUCATION



Mid-continent Research for Education and Learning (McREL) is a private, 501 (c)(3) education research and development corporation. At McREL's headquarters in Denver, Colorado, and McREL's Pacific Center for Changing the Odds in Honolulu, Hawaii, you'll find experienced researchers and education consultants working together to provide educators with research-based, practical guidance on the issues and challenges facing K-16 education.

In January 2011, McREL formed The Network for Innovative Education™ as a call to action for educational systems in the United States to once again become the "best in the world." We feel an urgency to call attention to the "other achievement gap" in our country—the one between our highest performing systems, schools, and students, and those in other countries. This *Noteworthy* tells the story of our beginning work.

Contact McREL to order additional copies of
Noteworthy Perspectives: High Reliability Organizations in Education at:

Mid-continent Research for Education and Learning
4601 DTC Blvd., Suite 500
Denver, CO 80237
P: 303.337.0990 • 800.781.0156 • F: 303.337.3005
E-mail: info@mcrel.org • website: www.mcrel.org

Noteworthy

P E R S P E C T I V E S

High Reliability Organizations in Education



Table of Contents

Chapter One: Introduction: Becoming the “Best in the World “at Educating Our Students	1
Chapter Two: Toward Highly Reliable, High-Quality Public Schooling	6
Chapter Three: High Reliability & Leadership for Educational Change	24
Chapter Four: Best in the World: High Performance with High Reliability	36

Chapter One: Introduction

Becoming the Best in the World at Educating Our Students

By James H. Eck, McREL

Once at the forefront of educational achievement by any number of measures, the United States has dropped in standing relative to its economic partners and competitors in the global marketplace. While we have taken incremental steps to improve student learning, other countries have made tremendous strides, not just catching up to us, but passing us by. As a result, U.S. students are now grossly under-represented in the ranks of top scorers on international tests, putting America's ability to maintain its position as a leader in innovation, technology, and prosperity at risk.

The Programme for International Student Assessment (PISA) exam is a battery of assessments of 15-year-old students in mathematics, science, and reading, administered internationally every three years. In 2006, the United States ranked 25th of 30 nations in mathematics and 24th of 30 in science (reading literacy was on an off-year). This puts us on par with Spain, Portugal, and the Slovak Republic, and far behind Canada, the Netherlands, Australia, Korea, and other countries. This, in spite of the pledge of the National Governors Association in 1989 that U.S. students would lead the world in mathematics and science achievement by 2000 (Walberg, 2003).

The United States has participated in all of the international tests since 1964 and has maintained a longitudinal testing system of its own, the National Assessment of Educational Progress (NAEP).

Stanford economist Eric Hanushek and his colleagues (Hanushek, Jamison, Jamison, & Woessmann, 2008) conducted a cross-assessment analysis of a variety of international tests in reading, mathematics, and science administered between 1964 and 2003 and calibrated each of the separate international tests to the NAEP. They found that performance of U.S. students over the past four decades remained flat (along with Germany and Hungary), while those of students in the Netherlands and Finland have improved. Asian countries have consistently performed well, including those that have entered the international assessment arena along the way. Poland was the most rapidly improving nation, registering average gains on the PISA of more than 25 points between 2000 and 2006 alone and surpassing the U.S. performance in 2006. (Hanushek, et al., 2008; McKinsey & Company, 2009; Organisation for Economic Co-operation and Development [OECD], 2010).

A series of research studies has led Mid-continent Research for Education and Learning (McREL) to an important conclusion about how to improve the international competitiveness of U.S. schools. As in the field of medicine, where some advances in ensuring better end results for patients have come not from technology or pharmaceuticals, but from simple improvements, including better hygiene in hospitals and the systematic use of checklists,

Top systems in the world

- * Alberta, Canada
- * Australia
- * Belgium
- * Finland
- * Hong Kong
- * Japan
- * Netherlands
- * New Zealand
- * Ontario
- * Singapore
- * South Korea

what we most need now in U.S. education is not more funding, more technology, or radical new innovations. Rather, the solution to ensuring America's students are able to compete on a global stage lies in re-thinking the role of schools and school districts. By assisting school systems to more closely resemble "high reliability" organizations (HROs) that already exist in other industries and benchmarking against top-performing education systems from around the globe, America's school systems can transform themselves from compliance-driven bureaucracies to world-class organizations.

Missing the bar

Research carried out by consultancy firm McKinsey & Company between May 2006 and March 2007 resulted in the report, *How the World's Best-*

Performing School Systems Come Out On Top (2007), in which the authors identify the highest performing school systems by their 2006 performance on the PISA exam. In addition, the McKinsey researchers had conducted extensive site visitations and included a comprehensive review of current literature, and interviews with experts, policymakers, and practitioners.

Blatantly missing from this list of top performers are school districts from the United States, although several (Atlanta, Boston, Chicago, and New York City, along with one state—Ohio) were identified as being on “strong improvement trajectories.” From an economic standpoint, which was one impetus for conducting such a study in the first place, the results are discouraging. Despite huge increases in spending for education and ambitious reform efforts, Americans are seeing little improvement across their school systems. Few of the most widely supported reform strategies (e.g., giving schools more autonomy, reducing class sizes) have produced the promised results (McKinsey & Company, 2007).

However, while looking at whether differences at the system level impacted student achievement by enabling better teaching and greater learning, McKinsey researchers found that the highest performing systems in the world, despite possessing large differences in culture, context, and construct, maintained a primary focus on instructional quality. These systems emphasized three things: (1) getting the best candidates into the teaching profession; (2) providing continuous, embedded in-service professional development; and (3) ensuring that the system responds to early signs of individual student failure.

Andreas Schleicher, head of the indicators and analysis division of the Organisation for Economic Co-operation and Development

(OECD), which administers the PISA exam, drew an interesting conclusion after examining the McKinsey & Company findings. He observed that the high-performing systems shared a relentless focus on ensuring high instructional quality, while at the same time, reducing variability in the instruction every child receives (Schleicher, 2008). At McREL, we are particularly concerned about variability in system performance within schools, among schools and districts in the United States, and between U.S. educational systems and the rest of world.

Aiming for high quality, low variability

Most of the attention to achievement gaps in the United States has focused on the persistent performance differences among subgroups of students by race/ethnicity and socioeconomic status, particularly in our urban school districts. Increasing diversity in subgroup populations in suburban and rural districts has also contributed to achievement gaps being exposed that may not have been present before, or that were attributed previously to just a few students and overlooked. Additionally, as data collection and reporting systems have improved, districts and schools that were previously considered “high performing” began to find achievement gaps between subgroups of students that may have been previously masked.

Yet, in the report, *The Economic Impact of the Achievement Gap in America's Schools*, McKinsey & Company (2009) stress the importance of looking at two other gaps: (1) between similar students schooled in different systems or regions of the country; and (2) between the United States and other nations. In fact, “the most striking, poorly understood, and ultimately hopeful fact about the educational achievement gaps in the United States

involves the huge differences in performance found between school systems, especially between systems serving similar students” (p. 12).

Possibly even more striking is the variability in instruction within schools. Hattie (2009) reports that, using multi-level modeling, researcher Spyros Konstantopoulos found a substantial proportion of the variation in student achievement lies within schools and not between schools. Grodsky and Gameron (as cited in Hattie, 2009) conclude that many of the influences that really make a difference to student learning in developed nations are within schools, from the influence of specific teachers, specific curriculum, and strategies teachers use to teach. Thus, one solution seems to be improving instructional quality while reducing the variability in the quality of that instruction within and among schools.

Defining system-level leadership and supports

Leadership plays a critical role in the performance of these “best in the world” systems. Indeed, the McKinsey & Company (2007) study notes that the research on school leadership suggests “school leadership is second only to classroom teaching as an influence on learning” (p. 29). Furthermore, they assert that school reforms rarely succeed without effective leadership, both at the level of the system, and at the level of individual schools. Researchers of another study noted that “there is not a single documented case of a school successfully turning around its pupil trajectory in the absence of talented leadership.” (Leithwood et al., 2006, p. 5).

How do we define a “system of education”?

The majority of the OECD countries compared in the McKinsey & Company report have nationalized

education. In the United States, the school district is the legally defined entity for public education, but the devolution of “local control” varies from state to state. State departments of education and the U.S. Department of Education represent additional levels of scale, but they have incrementally less decision-making authority (although that has tightened substantially in the past three decades). Therefore, as we think about systemic solutions for system-level issues, we will focus primarily on the district and then look one level down to the school and up two levels to the state and national levels.

In the book, *District Leadership that Works: Striking the Right Balance*, Marzano and Waters (2009) present five district-level responsibilities from their meta-analysis that were found to be statistically correlated with increased student achievement. The research initially set out to answer the question of whether superintendent leadership in a district had an effect on student achievement. The findings from the study, however, indicate the importance of an expanded definition of district-level leadership, to include in addition to the superintendent, the collective central office staff, the board of education, and principals with their schools, operating as a school district versus a district of schools. The central theme across the responsibilities is the need to establish non-negotiable goals for achievement and instruction across the district as a whole.

Emulating High Reliability Organizations

Marzano and Waters (2009) went on to consider their findings about district leadership and defined autonomy from the perspective of High Reliability Organizations. Karl Weick has focused his research agenda in the last three decades to examining these organizations that “operate under high risk conditions

and take a variety of steps in pursuit of error-free performance” (Weick, Sutcliffe, & Obstfeld, 1999). Weick’s work with HROs began with studying flight deck operations on a nuclear aircraft carrier and has carried over to research including nuclear power plants, wildland firefighting, and aircraft flight operations.

In these industries, any mistake can have disastrous consequences—people die. To avoid disaster, these organizations put into place multilayered structures and processes to prevent errors and more importantly, to respond quickly before errors can cascade into catastrophic system failures. They also *mindfully* anticipate and manage the unexpected (Weick & Sutcliffe, 2001, 2007; Hoy & Sweetland, 2001; Hoy, 2003). Errors and mistakes are bound to occur, but the key is anticipating that they will occur and responding to them as soon as they appear. This constant monitoring for the early signs of failure and responding quickly is another way HROs demonstrate the characteristic of mindfulness.

At this point, a logical question to ask is, “What do these organizations have in common with K–12 public education systems and what can we possibly learn from them?” At McREL, we are translating the McKinsey & Company findings from the world’s highest performing educational systems through a lens of high reliability. An operational definition of high reliability, applied to these systems, is this: *high levels of student performance, achieved as a result of high-quality instruction, delivered through superior execution of effective research-based practices, with low variability in the quality of instruction within and between schools.*

As we started to explore this idea of higher reliability educational systems, we ran across two other sets of educational researchers who were doing similar work. Sam Stringfield

and David Reynolds began their theoretical exploration of HROs in 1991 (Stringfield, 1991) and, with the addition of Gene Schaffer, initiated a set of High-Reliability Systems (HRS) research studies in Great Britain in 1995. Stringfield and several of his colleagues have a long research history on the topics of teacher and school effectiveness and system improvement (Stringfield, 1991; Teddlie & Stringfield, 1993; Stringfield, Millsap, & Herman, 1998; Reynolds, Creemers, Stringfield, Teddlie, & Schaffer, 2002). Stringfield, Reynolds, and Schaffer approached their HRS project from an assumption that practices gleaned from these fields could be coupled with HRO concepts to establish a school improvement strategy.

Tom Bellamy and his colleagues also were examining the topic and wrote the article, “The Fail-Safe Schools Challenge: Leadership Possibilities for High Reliability Organizations” (Bellamy, Crawford, Huber-Marshall, & Coulter, 2005), in which they presented HRO as at least a metaphor, if not a model for education. They asserted:

The stakes for failure have been raised so high ... that high reliability has become an important aspect of school success. Schools are now challenged to prevent practically all failures and to close achievement (gaps) among student groups—in short, to ensure highly reliable learning for all students.” (p. 384)

The hidden cost of underperformance

Although much attention remains focused on student failure (and rightly so), the ability of the United States to remain a global leader in innovation, science, technology, patents conferred, business, and social entrepreneurship will depend on the ability of its educational systems to

not only raise the floor, but also the ceiling. There are certainly moral and ethical aspects for closing the variety of achievement gaps, and severe social implications if we do not. Conversely, the economic benefits of dramatically raising the bar for the U.S. education system can be tremendous.

Economist Eric Hanushek's research has been not only on calibrating international assessments to one another for comparative performance, but also in constructing a sophisticated methodology for linking cognitive performance to economic growth, in terms of Gross Domestic Product (Hanushek, et al., 2008; Hanushek, Peterson, & Woessmann, 2010). Hanushek and his colleagues calculate if the United States had closed the gap between its educational achievement levels and those of other countries such as Korea and Finland, 2008 Gross Domestic Product could have been \$1.3 trillion to \$2.3 trillion higher (9%–16% of GDP). Furthermore, some see the persistence of these educational achievement gaps as impacting the U.S. economy with the equivalent of a permanent national recession (McKinsey & Company, 2009).

A case for urgency, a call for action

A move to high-performing, high-reliability, failure-free schools will, of course, require changes in cultures and systems. It will require that educators, policymakers, and the public examine basic assumptions about education in the United States and learn from the high performers, both here and abroad, to make America's schools among the best in the world.

On October 27–28, 2010, McREL, with support from the Kern Family Foundation, convened a small group of international thought leaders,

forward-thinking superintendents, CEOs from educational organizations, and leaders from high-performance professions. Together, we explored the “new frontier” in improving the performance of U.S. schools and districts—a frontier that lies not in dreaming up new innovations or more “silver bullet” fixes for education, but rather, in flawless implementation of existing know-how to ensure all students benefit from top-quality instruction and learning environments.

At the Best in the World (BITW)¹ gathering, we particularly focused our attention on “the other achievement gap,” the difference in performance between America's educational systems (even our highest performing districts) and those among the best in the world. The stated premises for this gathering were these:

- There is a gap in achievement between America's highest performing schools and school districts and the highest performing systems internationally.
- This gap in achievement may be a more serious threat to the future of the country than the gap between high-performing and low-performing U.S. schools and districts.
- This gap in achievement can only be closed by “raising the ceiling,” or elevating the performance of America's highest performing schools and school districts.
- Elevating system performance, without excluding large numbers of students, requires a commitment to high performance with high reliability (raising the ceiling and the floor).
- Creating a constituency for, and urgency about, high-performing, high-reliability schools and

districts is the biggest challenge we face in U.S. education.

The presentations from this group of experts (see sidebar on p. 5 for list of presenters) cemented the case that much can be learned from international comparisons, both from international comparative measures, such as PISA, and from benchmarking what works from those systems to school districts in the United States.

We understand the urgency, and we know that throwing large sums of money and a barrage of reform efforts at the problem hasn't resulted in significant, sustainable, or scalable change, but what is the best thing to do?

A theory of action—high performance with high reliability

While our focus for this gathering was on “the other achievement gap” between the United States and other systems of education worldwide, in light of Schleicher's conclusion, the variability in achievement found among, and within, state and district educational systems across the United States requires simultaneous attention. We believe that lessons learned from High Reliability Organizations may provide us with a foundation for school improvement and with a set of principles and strategies to directly apply to our educational systems. We asked Stringfield and his colleagues along with Tom Bellamy, to help us develop a theory of action.

Bellamy and Stringfield also prepared commissioned papers to accompany their presentations at the Best in World Exploratory Gathering. Those two papers constitute the next two chapters of this monograph. We invite you to read on.

¹ For video clips of the presentations, visit the Network for Innovative Education website at <https://sites.google.com/site/networkforinnovativeeducation/Home>.

Best in the World presenters

- * Sir Michael Barber, expert partner in McKinsey & Company's Global Public Sector Practice and head of Global Education Practice, and co-author of *How the World's Best-Performing School Systems Come Out On Top*.
- * Martin West, deputy director of the Program on Education Policy and Governance, Harvard University
- * Tom Bellamy, professor and director, Goodlad Institute for Educational Renewal, University of Washington Bothell
- * Sam Stringfield, professor and distinguished university scholar, University of Louisville; with Eugene Schaffer, professor and chair, Dept. of Education, University of Maryland Baltimore County; and David Reynolds, professor, University College Cardiff, Wales

References

- Bellamy, T., Crawford, L., Huber-Marshall, L., & Coulter, G. (2005). The fail-safe schools challenge: Leadership possibilities from high reliability organizations. *Educational Administration Quarterly*, 4(3), 383–412.
- Hanushek, E. A., Jamison, D. T., Jamison, E. A., & Woessmann, L. (2008, Spring). Education and economic growth. *Education Next*, 62–70.
- Hanushek, E. A., Peterson, P. E., & Woessmann, L. (2010, November). *U.S. math performance in global perspective: How well does each state do at producing high achieving students?* Retrieved from http://www.hks.harvard.edu/pepg/PDF/Papers/PEPG10-19_HanushekPetersonWoessmann.pdf
- Hattie, J. A. C. (2009). *Visible learning: A synthesis of meta-analyses relating to achievement*. New York, NY: Routledge.
- Hoy, W. K. (2003). An analysis of enabling and mindful school structures: Some theoretical research and practical considerations. *Educational Administration Quarterly*, 41(1), 87–108.
- Hoy, W. K., & Sweetland, S. R. (2001). Designing better schools: The meaning and measure of enabling school structures. *Educational Administration Quarterly*, 37(3), 296–321.
- Leithwood, K., Day, C., Sammons, P., Harris, A., & Hopkins, D. (2006). *Seven strong claims about successful school leadership*. London: Department for Education and Skills.
- Marzano, R., & Waters, T. (2009). *District leadership that works*. Bloomington, IN: Solution Tree.
- McKinsey & Company. (2007). *How the world's best performing school systems come out on top*. New York, NY: Author.
- McKinsey & Company, Social Sector Office. (2009, April). *The economic impact of the achievement gap in America's schools*. Retrieved from http://www.mckinsey.com/app_media/images/page_images/offices/socialsector/pdf/achievement_gap_report.pdf
- OECD Programme for International Student Assessment (PISA). (2010). *The high cost of low educational performance*. Paris, France: Author.
- Reynolds, D., Creemets, B., Stringfield, S., Teddlie, C., & Schaffer, E. (2002). *World class schools: International perspectives on school effectiveness*. New York, NY: Routledge/Falmer.
- Schleicher, A. (2008, June). *Seeing the U.S. school systems through the prism of international benchmarks*. Paper presented at the Annual Policy and Practice Institute for School Leadership, Dover, DE.
- Stringfield, S. (1991, January). *A model of elementary schooling effects: Evidence of support*. Paper presented at the meeting of the International Congress for School Effectiveness and Improvement, Cardiff, Wales.
- Stringfield, S., Millsap, M., & Herman, R. (1998). Using "promising programs" to improve educational processes and student outcomes. In A. Hargreaves, A. Lieberman, M. Fullan, & D. Hopkins (Eds.), *International Handbook of Educational Change* (pp. 1314–1338). Dordrecht, The Netherlands: Kluwer.
- Teddlie, C., & Stringfield, S. (1993). *Schools make a difference*. New York, NY: Teachers College Press.
- Walberg, H. J. (2003, Spring). Accountability unplugged. *Education Next*, 76–79.
- Weick, K. E., & Sutcliffe, K. M. (2001). *Managing the unexpected: Assuring high performance in an age of complexity*. San Francisco, CA: Jossey-Bass.
- Weick, K. E., & Sutcliffe, K. M. (2007). *Managing the unexpected: Resilient performance in an age of uncertainty*. San Francisco, CA: Jossey-Bass.
- Weick, K. E., Sutcliffe, K., & Obstfeld, D. (1999). Organizing for high reliability: Process of collective mindfulness. *Research in Organizational Behavior*, 21, 81–123.

Chapter Two

Toward Highly Reliable, High-Quality Public Schooling

By Sam Stringfield, University of Louisville; David Reynolds, University of Southampton;
and Eugene Schaffer, University of Maryland, Baltimore County

Introduction

Large numbers of U.S. schools and their students can achieve at levels fully competitive with, and perhaps redefining, “the best in the world.” Furthermore, empirical evidence suggests that dramatic improvement in American educational reliability—educating virtually all children to fundamentally higher levels of measured achievement—is possible.

The challenge of providing “the best in the world” education involves knowing “what works extremely well” and providing it with remarkable reliability. These two components are multiplicative; that is to say, the quality of education received by students in any given classroom, school, local education authority (LEA), state, or nation is the product of the extent to which the providers use state-of-the-art, proven methods and processes multiplied by the reliability of delivery. This idea, stated as a straightforward equation, becomes this:

Quality of Education = Extent of Methods × Reliability of Delivery

Much more—and typically more rigorous—research exists on the subject of “what can be effective” than on “how to reliably deliver it.” To partially re-balance this situation, our focus in this chapter is on

methods for improving the reliability of educational reform efforts.

Trend data and implications

High Reliable Organizations evolve only when the professionals working in an area and the larger public come to believe that the historic levels of the organization’s reliability are likely to lead to disaster. The data we examined and present here indicate that the United States is moving precisely toward that condition. We observe, too, that in such complex systems, teachers and other educators see a “loosely coupled system” (Weick, 1976) and respond, for purposes of their long-term professional survival, as “street level bureaucrats” (Lipsky, 1980). Stated in more theoretical terms, when fundamentally interesting, potentially valid reforms have been attempted in parts of the complex education system, they were predicted to fail—and often did—not because the ideas were invalid, but because they were overwhelmed by the larger logic of the system. A point to which we will return repeatedly is that in educational reform, as in research, reliability sets the upper boundary of measured validity.

International trend data

The Trends in International Mathematics and Science Study (TIMSS)¹ provides an unparalleled

set of studies spanning more than a decade and providing comparisons and contrasts of student achievement among nations’ systems of schooling. While lacking the breadth of TIMSS, Reynolds, Creemers, Stringfield, Teddlie, and Schaffer (2002) produced a mixed-methods study contrasting higher and lower achieving schools within and across nine countries’ educational systems. Both of these studies examine the comparable value of schooling on student test performance. The first looks at cross-national data while the second looks at within-country variance and cross-country variance following cohorts for two years.

The National Center for Educational Statistics (NCES) regularly provides reviews of various international comparisons of student achievement (TIMSS, the Progress in International Reading Literacy Study [PIRLS], and the Program for International Student Assessment [PISA]). NCES reported that in 4th- and 8th-grade reading, U.S. students ranked 10th of 45 nations/provinces studied. Disturbingly, the NCES review found that over time, a growing number of countries’ students were exceeding the average of U.S. students in reading abilities. In 4th- and 8th-grade mathematics, the TIMSS data indicate that U.S. children are making progress over time relative to students in other countries, and

¹ For an overview and range of reports, see <http://nces.ed.gov/timss>.

currently rank in the top 8 of over 40 countries on mathematics measures.² In the 2006 PISA science study, U.S. 15-year-olds scored in the bottom third of OECD (economically developed) participating nations. To the extent that our goal is to be “the best in the world,” we have some distance to travel.

U.S. longitudinal achievement trends

Complaints about “the current state of schooling” in the United States and the need for “dramatic improvements” have been staples of the American political scene for over 200 years. Consider that at the end of the 19th century, Harvard’s president complained that the American students at college entry simply were not up to European standards. Eliot (1898) laid the blame squarely on American schools, whose “inain characteristic of instruction is dullness, a complete lack of human interest and a consequent lack in the child of the sense of increasing power” (p. 184, as cited in Nunnery, 1998). Nearly a century later, John Goodlad (1984), summarizing a large study he and a team had completed, observed, “Only rarely did we find evidence to suggest instruction likely to go much beyond mere possession of information... Boredom is a disease of epidemic proportions” (pp. 236–242).

We posit the sameness of the critiques is not the result of lack of change efforts and offer a few facts regarding educational outcomes over the last century, the relative stability over the last 30 years, and the costs of that sameness in an ever-changing and educationally improving world:

1. The overall percentages of Americans per birth cohort who graduated from high school stood

at approximately 20 percent at the dawn of the 20th century, and rose to 76 percent in 1970. The high school graduation rate then gradually declined to 68 percent in 1998, and subsequently has risen steadily to its current 75–77 percent today. The good news here is that in the first decade of the

21st century, the rise in the percent of students graduating from high school is as steep as at any time in our nation’s recorded history (Heckman & LaFontaine, 2010). The bad news is that a quarter of our young people are leaving education almost completely unprepared to compete in a global 21st century information economy.

2. Of our non-high school graduates, the major change has been in the percentage taking and passing the General Educational Development tests (GEDs), which is considered to be a high school equivalency certificate. The percentage of high school dropouts taking the GED tests has increased rapidly over the past 15 years and helps explain the substantial gaps between U.S. Census data on percentages of young adults who are “high school graduates” (including, for Census purposes, GED holders) and data on actual graduation rates. Unfortunately, today the GED provides very modest economic value over high school dropouts not completing a GED. By contrast, high school graduation has approximately a 50 percent “value added” over not graduating.
3. The percentages of Americans with four-year college degrees or higher has risen steadily throughout the last century, from approximately 5 percent of the cohort born in 1900

to approximately 31 percent today. Interestingly, the nation’s gains over the last 30 years in percentages of young people graduating from college have been the result of increasing percentages of high school graduates—especially females—attending and completing college.

4. For nearly four decades, the U.S. Department of Education has conducted extensive, nationally representative studies of student achievement in the areas of reading and mathematics. These data comprise the National Assessment of Educational Progress (NAEP) long-term trend data.³ Figure 1 provides data, presented in mean scale scores, on NAEP student reading scores at ages 9, 13, and 17 from 1971 through 2008. The relatively good news in Figure 1 is that age 9 reading mean scores are at the highest level yet measured on NAEP. The less encouraging news is that the age 13 scores are at the same level as in 1992, and are not dramatically above the levels of 1971. The least encouraging news is that the mean score for 17-year-olds is not statistically different from the scores in 1971 and is actually significantly below the scores from the mid-1980s through early 1990s.
5. Figure 2 provides NAEP mean mathematics scores for 9-, 13-, and 17-year-olds from 1973 through 2008. Again, the most encouraging news is from the 9-year-olds. Those scores have risen significantly and at an impressive rate of progress over the last 35 years, with the steepest rise coming in the last decade. Statistically significant and only moderately

² See <http://nces.ed.gov/programs/coe/2009/analysis>.

³ See <http://nces.ed.gov/nationsreportcard/lt>.

Figure 1

Trend in NAEP Reading average scores for 9-, 13-, and 17-year-old students, 1971–2008

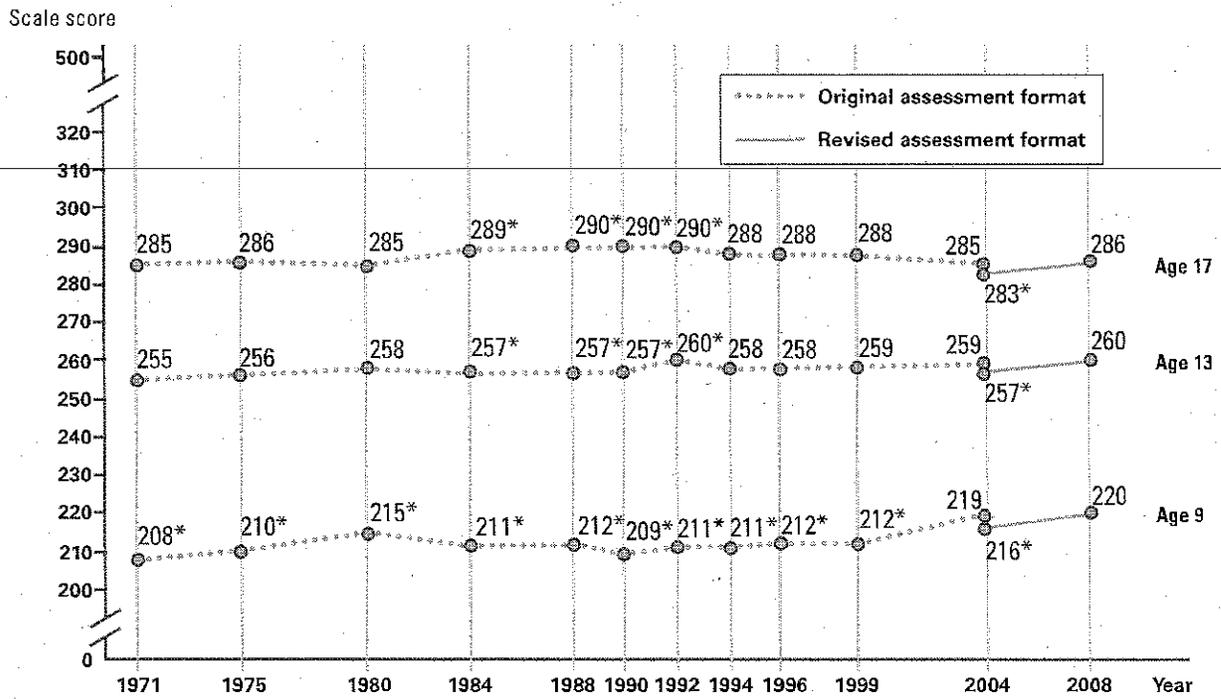
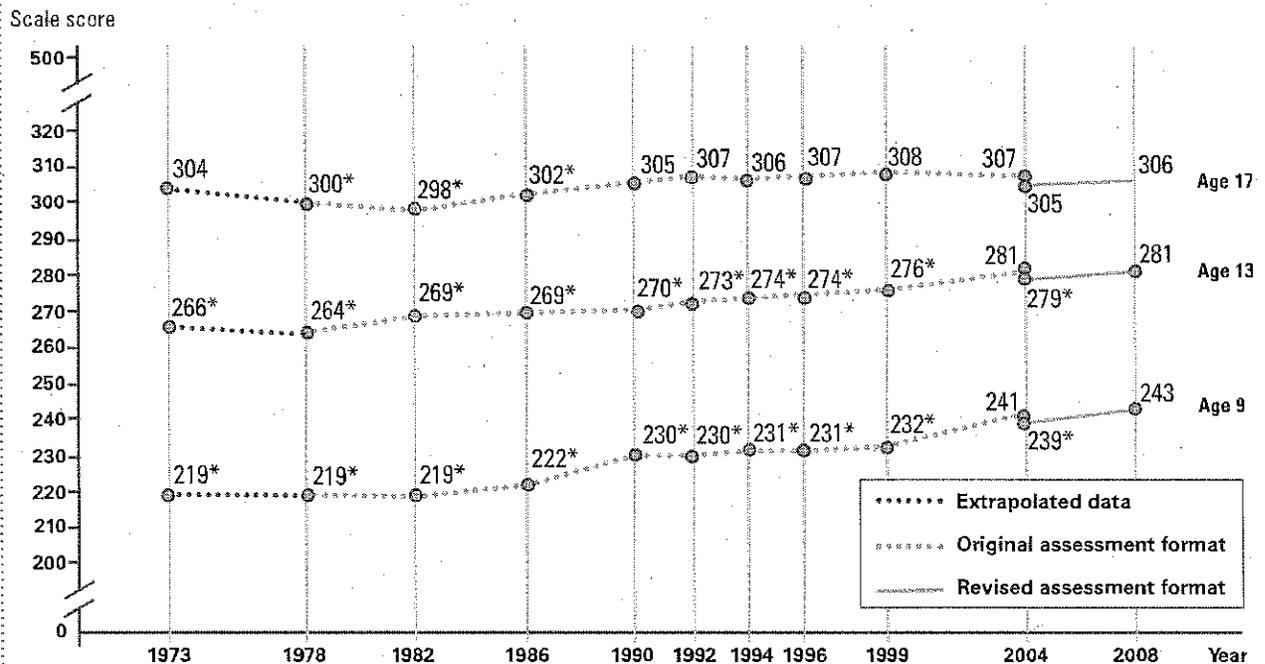


Figure 2

Trend in NAEP Mathematics average scores for 9-, 13-, and 17-year-old students, 1973–2008



less impressive progress has been made by our schools in advancing the mathematics achievements of 13-year-olds. However, as with reading scores, 17-year-olds' mean math scores in 2008 were not statistically different from scores in the early 1970s and do not reflect a significant rise in the last decade.

In summary, our level of high school graduation has risen for most of the last century, fell modestly for two decades, and is again at its highest levels in history, at approximately 75 percent. Including holders of GED certificates, over 85 percent of young adult Americans are counted as being "high school graduates." Rates of college attendance and graduation have risen almost continuously through the last century and today stand at 31 percent of each new birth cohort. The best available evidence of long-term reading and mathematics achievement among American school-aged students shows clear 30+ year gains in the elementary grades, but no appreciable gains by the upper grades of high school, as students prepare for college and careers, or both. So, what does this mean, exactly?

Economic and other impacts of education in the United States

Figure 3 presents data from 1949–2000 on the median income of young adult (age 25–34) American males. We focus on young adults because the effects of education are first noticeable in the age range that could be expected to have completed their formal education. We focus on males because in the late 1940s only 25 percent of U.S. females were working outside the home, and that number had risen to 80 percent by the year 2000, hence making whole-cohort comparisons among females over time problematic. One of the most striking sets of facts revealed in Figure 3 is the change in the economic benefit

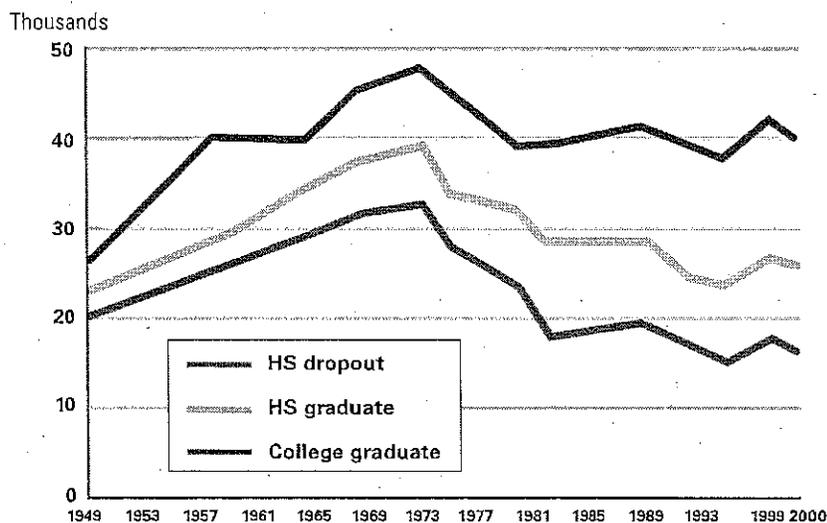
of obtaining additional education. In 1949, the average young male college graduate made 30 percent more in annual income than the average high school dropout. (In other words, for every dollar the average young male high school dropout earned, a college graduate earned \$1.30.) By the year 2000, that differential had expanded to over 150 percent. (For every dollar earned by a young male dropout, the college graduate made over \$2.50.) The economic advantage of succeeding in schooling had increased by over five-fold. In constant (inflation-adjusted) dollars, a 21st century high school dropout not only makes less than his grandfather, the high school dropout, made after World War II, he makes less than half as much as his father, the high school dropout, made in the early 1970s.

For a current picture, we need not limit the discussion to males or young people. Figure 4 presents a point-in-time data set of all Americans' (male and female, all ages) income

by education in 2008. According to the U.S. Bureau of Labor Statistics, across all persons in the would-be-working range of our population, the probability of not being able to find gainful employment was over three times higher for high school dropouts than for college graduates, regardless of age or gender. Regarding income, for every dollar the average high school dropout earned, the average high school graduate made \$1.38, the average college graduate made \$2.30, and the average professional (Ph.D., M.D., etc.) made over \$3.50. Discrepancies of this magnitude were almost unimaginable in the late 1940s and 1950s, but they are today's realities. Further expanding the differences, the typical college graduate marries another college graduate, with the practical implication being that the differences in family income by education often are doubled. Finally, in 2010, for the first time in U.S. history, more women are working outside the home than

Figure 3

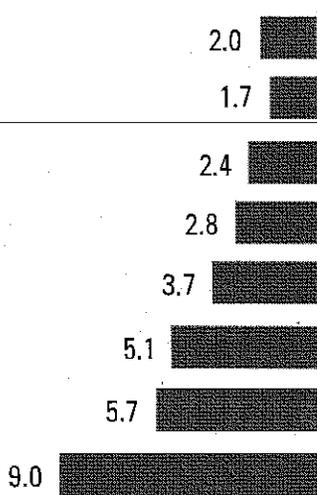
Median income of male Americans aged 25–34, by educational level, 1949–2000



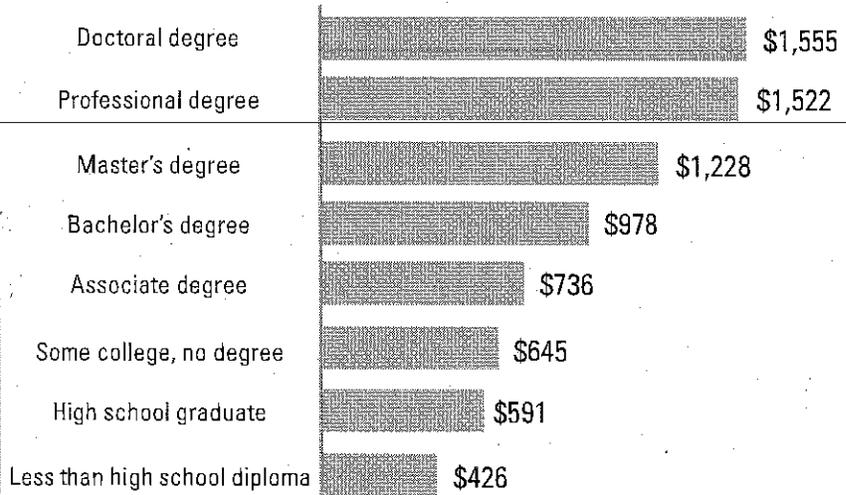
Source: Educational attainment in the United States, U.S. Census Bureau, March 2000. 2000 dollars (CPI-U adjusted)

Figure 4
Education pays

Unemployment rate in 2008



Median weekly earnings in 2008



Source: Bureau of Labor Statistics, Current Population Survey

Note: The BLS unemployment statistic averaged 6.2% in 2008. In September of 2010 it stood at 9.7%.

men, and more women are working in managerial positions.

We do not believe that income and employment are the most important variables in life, only that the longest span of reasonably reliable data are available on them. Regarding other variables, a range of studies have found that persons with higher levels of educational attainment are more likely to engage in a wide series of additional, pro-social activities. Whether the topic is voting, taking leadership positions in organizations ranging from churches to political parties, starting new businesses or staying off welfare and out of prison, increases in education are consistently correlated with success in life. The conclusion of studies of income and of a range of other desirable adult outcomes must be that success in school now matters more than at any other time in our history. Today, educational failure is catastrophic for the individual, his or her future family,

their community, and our society. So, why is it that seemingly logically valuable reform efforts have tended to not produce positive results? Complex, loosely coupled systems, and the inevitable roles of "street-level bureaucrats" Education in the United States is a very complex, loosely coupled system. This became abundantly clear to Stringfield when he was appointed to the New Board of School Commissioners of the Baltimore City Public Schools System (BCPSS) in 1998. He soon found the challenges bewildering (see Stringfield & Yakimowski, 2005; Stringfield, 2008). A particular point of importance—and frustration—in his experience was the search for credible levers for affecting change from the seemingly powerful position of the school board. This led to a series of attempts to model the relationships among various levels of the modern American education system (Datnow,

Lasky, Stringfield, & Teddlie, 2006; Land & Stringfield, 2005). No shortage of examples of complex educational systems Figure 5 (see p. 12) presents Stringfield's (2005) representation of the complex set of relationships among levels of educational governance in the United States. To take one practical example of these relationships, consider the passage of the federal No Child Left Behind (NCLB, 2001) legislation. Congress passed the legislation, but the U.S. Department of Education took over a year to develop regulations and even longer to develop non-regulatory guidance. (With federal educational legislation, states more often look to the non-regulatory guidance to determine how to meet legal requirements.) States then developed new testing schemes and established new regulatory requirements of their own, which they passed on to LEAs.

Meanwhile, colleges of education across the country were changing requirements in various courses and developing new programs to assist schools and districts in meeting the new requirements. More aggressively, a broad range of for-profit corporations (such as text and software publishers and consulting firms) and not-for-profit entities (such as the regional laboratories and various foundations) began developing products, workshops, and other materials to assist schools and LEAs in addressing the changes required in NCLB. LEAs received this range of information and federal funds, and made diverse new requirements on schools, which in turn made new demands on teachers. The theory of action here is that federal laws, coupled with funding that totals less than six percent of the average districts' budgets, will produce substantial change at each subsequent level, eventually resulting in increased student achievement. At the level of Figure 5, this presumes a tidy, tightly coupled system for educational improvement.

However, even the slightest examination of what happens within any one of the components of the system, let alone across the full set, tells a different story. Figure 6 follows the formal, official model of how changes happen within just one box of Figure 5. The one box is the LEA, but any other would make the same point. Within Figure 6 (see p. 13), a school board—typically elected, but in some large systems appointed—considers the policy implications of new laws and other inputs, sets policy, and directs the superintendent (the board's one employee) to implement policy. The superintendent works through the various offices under his or her control (deputy superintendents

for each school level, professional development, accounting, and testing, among others), and that group works with schools to implement the mandated-from-above changes. Goals are set and communicated, special programs are created or re-directed, and standardized tests and other measures provide feedback loops. Clear lines of authority exist within LEAs, and in theory, these are relatively tightly coupled.

Figure 7 (see p. 13) presents something closer to Stringfield's school board experience at the LEA level. The board did all of the things that the presumed theory of action dictated: considered options, established goals, directed the flow of money (while checking to be sure that the budget balanced), agreed on measures, and provided clear direction to the superintendent. The superintendent met with his district leadership team, discussed tactical options, passed down practical strategies, and so on. But almost immediately, a complex series of loops appeared.

Examples were everywhere. A board member's long-term neighbor and trusted friend (a teacher or a principal) came to the board member's home to complain bitterly. A third-level functionary somewhere in the central office discovered an inconsistency between new policies and old ones that had guided practical actions for 10 years. The reallocation of Title I funds was perceived by an elected official's spouse to disadvantage their 5th-grade child. A politically well-connected principal who had, over years, garnered great community support for her school either ignored the new directives with impunity or, if pushed, organized formal opposition and demanded change to what had

been. End-of-year test scores came in, and they did not demonstrate dramatic short-term effects of the by-now-unpopular changes. What quickly became apparent were the second-, third-, and fourth layers of formal and informal communications and powerful counter-veiling change forces.⁴ Almost immediately, the theoretically somewhat tightly coupled systems proved to be loosely coupled, calling for a different theory of action and related strategies for change.

"Loose coupling" is a term Karl Weick (1976) used to describe the working of schools, and we will return to it shortly. First, we briefly describe one aspect of educators' worlds, and their "street-level" policy implementation.

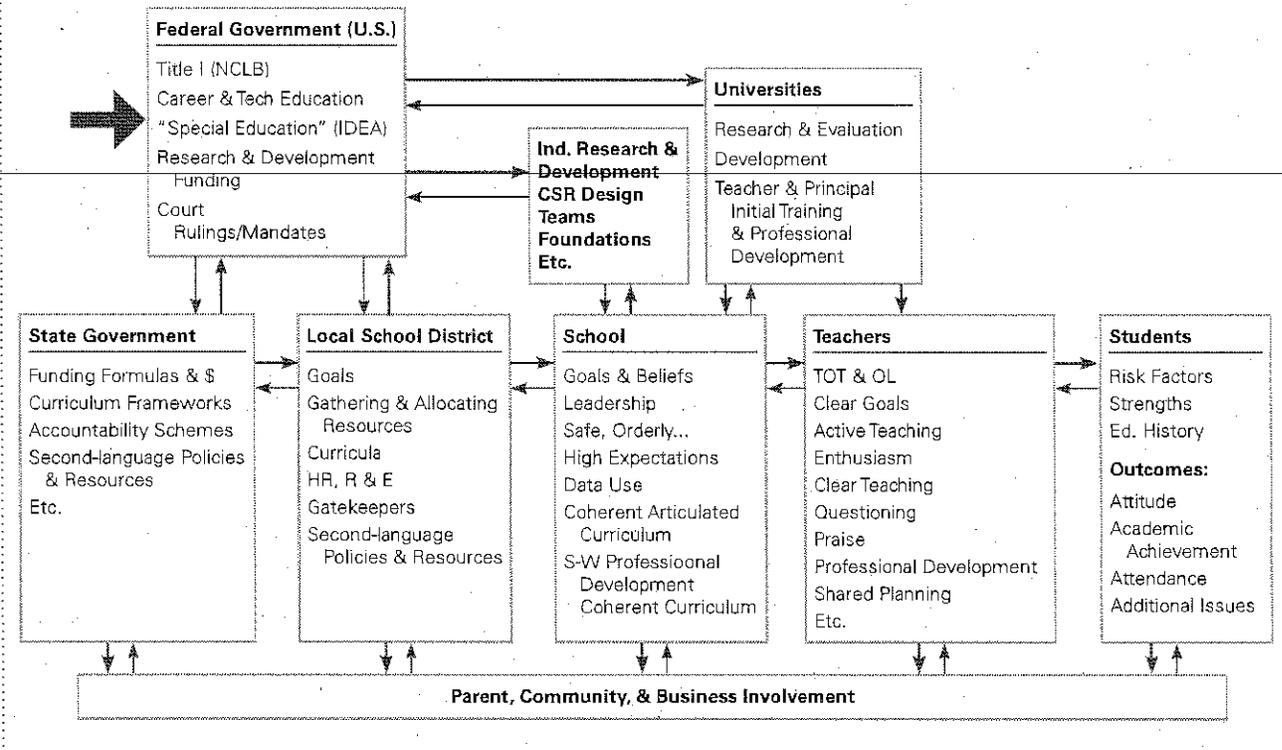
Street-level bureaucrats

Teachers, principals, and local educational administrators are necessarily practical people. Faced with simultaneous requirements to continue work in their classrooms (schools, district offices, etc.) and to implement changes, the full range of implications of which have almost necessarily not been thought through by those making the new demands, practical educators necessarily behave as "street-level bureaucrats" (Lipsky, 1980). Lipsky described autonomous workers, such as police officers and social workers, as working in arenas enmeshed in vague and often conflicting goals, accountability requirements, large demands for services—often more than an individual can provide, and often to involuntary clients—and the additional requirement of performing with limited and typically inadequate resources. These employees must find ways to manage under requirements that, if taken literally, would be impossible for any one human being

⁴As this chapter is being written, opponents of Louisville's (Ky.) long-standing student busing system are simultaneously in court arguing for a 100 percent return to neighborhood schools and mobilizing support to vote out board members who have supported the policy that once had been mandated by federal courts. Change forces are more complex than most of us realize until we try to change something.

Figure 5

A static representation of the relationships of educational organization levels and their potential influences on students



to implement. So, these street-level bureaucrats negotiate the space in ways they individually deem best. Policemen decide which of the thousands of laws to enforce as they walk their beat; social workers decide which clients to turn in for minor infractions, and when to look the other way.

Faced with requirements to maintain at least a minimum level of classroom discipline; collect lunch money; prepare lessons; write, administer, and score tests; and literally hundreds of other, frequently changing tasks, educators, like all other street-level bureaucrats, make choices as to which policies, old and new, to implement this hour, day, and year. Hardly surprisingly, over time they develop sophisticated personal and group systems for filtering and interpreting new requirements that may or may not be achievable when added to

current tasks. In short, educators at several levels work in very complex, often conflicting environments. They must interpret each new signal and decide whether they can respond, and if so, determine how best to do so. It is at this "street level" that educational change does or does not become a reality.

Loose coupling

Karl Weick (1976) observed that educational organizations are "loosely coupled systems" (p. 1) and noted several advantages. On the upside, loose coupling allows some portions of an organization to persist. Loose coupling prevents each part of an organization from having to respond to every single new signal in a system (i.e., no one has to do the impossible continuously).

In addition, loose coupling allows for localized adaptation. Persons and

groups face different challenges and are often best served by addressing their situations differently. These not-centrally-planned mutations sometimes eventually prove valuable to the larger whole, and certainly they can facilitate local functioning. (However, loose coupling makes it much harder for an organization to change as a whole unit.) If there is a breakdown in one portion of a loosely coupled system, the breakdown need not affect other parts of the system. Given that people enjoy holding a sense of self-control and self-efficacy, it is noteworthy that loose coupling generates autonomy. This enhances individuals' sense of self-determination and may raise morale.

Regarding cost, Weick hypothesized that loosely coupled systems "should be relatively inexpensive to run, because it takes time and money

Figure 6

The local school district

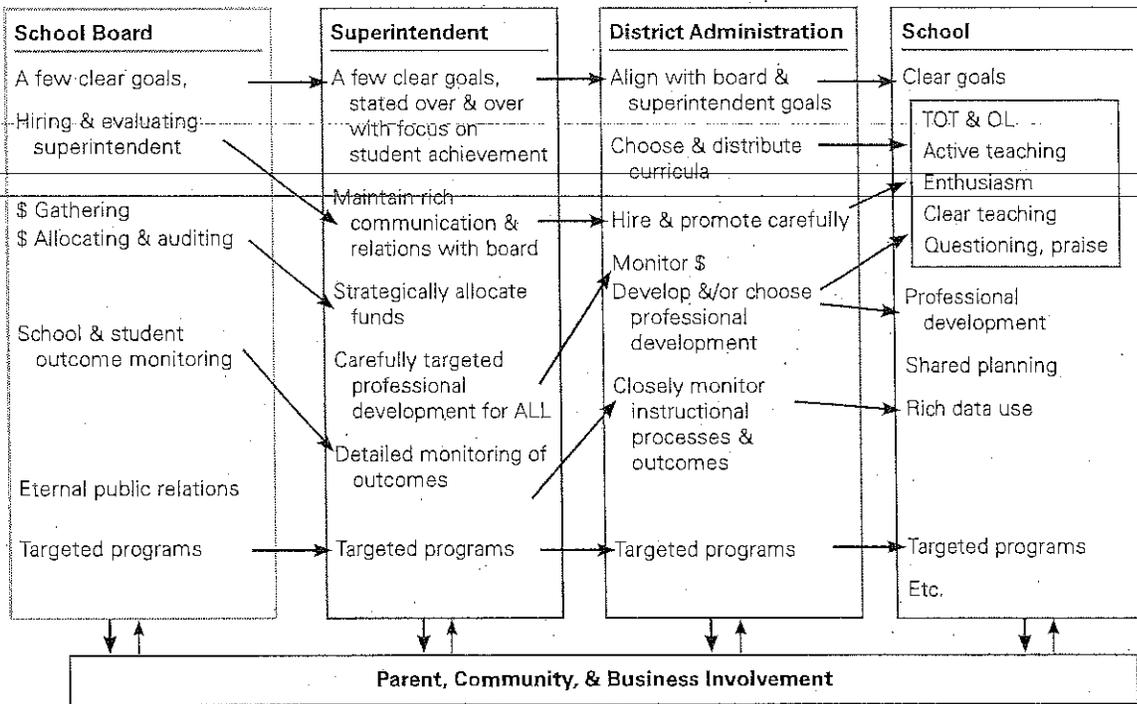
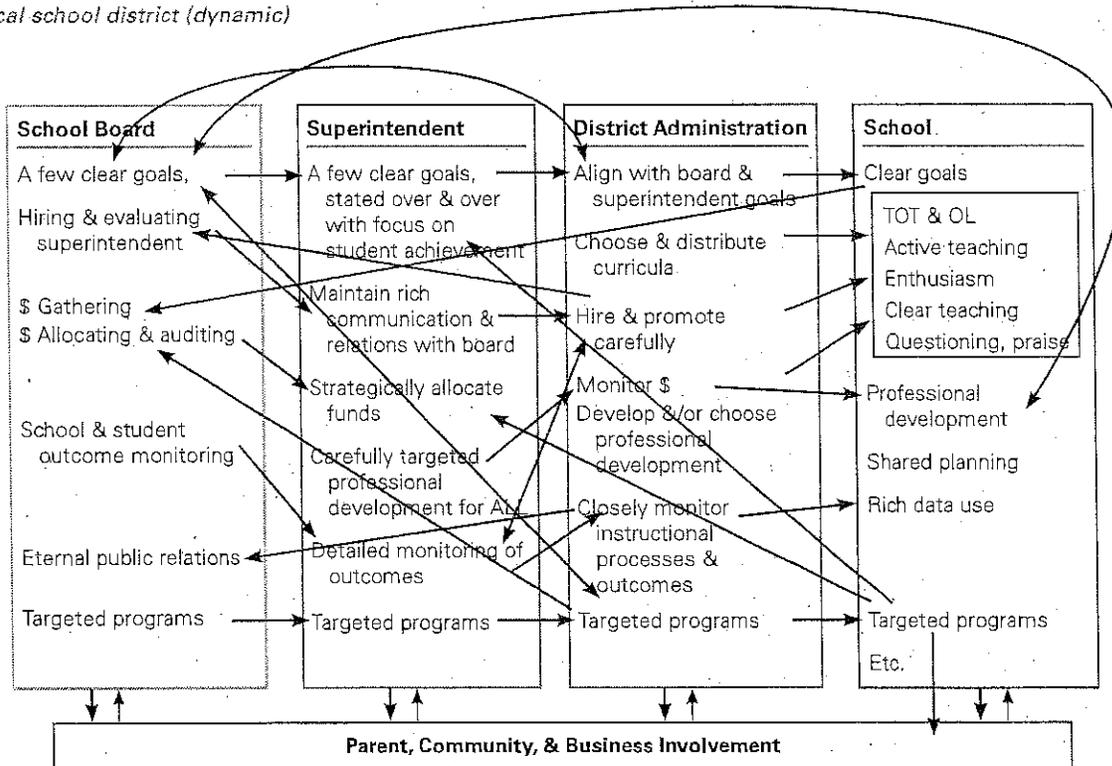


Figure 7

The local school district (dynamic)



to coordinate people” (p. 8). The tradeoff for being inexpensive is that loose coupling produces non-rational systems of fund allocation.

Finally, Weick observed that no organization could be tightly coupled in all areas. Tight coupling in some areas requires loose coupling in others. The ratio of tight and loose coupling varies among organizations.

Weick was neither exclusively pro- nor anti-loose coupling. His goal was to describe “what is.” An oft-expressed, street-level way of showing Weick’s central points is the image of “the egg carton school” in which each teacher may close the classroom door and do as she or he pleases in relative isolation and independence. The same egg carton model has historically described the relationships among schools and between schools and LEAs, LEAs and states, and the 50 state departments of education and the U.S. Department of Education.⁵

We believe that the combination of teachers, schools, and school systems working in loosely coupled relationships, and of educators necessarily making street-level decisions on an ongoing basis describes much of modern American educational practice. Among educators, there often is a firmly held attitude that research has little to nothing to offer practical educators⁶ and that most school systems have a new “focus” for each new school year, therefore ensuring the current focus will be gone next year, if not sooner. Loose coupling, when combined with these two commonly held beliefs, explains much of the failure of various educational reform

efforts to gain traction within and across schools.

Complex interventions inserted into inadequately understood, deeply complex, loosely coupled systems are unlikely to have measurable effects, even if they intermittently reach the street-level bureaucrats charged with implementing them. Under-funded, under-supported, inadequately trained street-level bureaucrats have few choices other than to keep moving to the next mandate. Given such realities, and remembering that reliability sets the upper boundary of measured validity, nearly all of us would predict that almost all reform efforts would fail.

Serendipity favors the prepared minds: The evolution of High Reliability Schools (HRS)

In the summer of 1989, Stringfield was trying to solve a data riddle. He and Charles Teddlie had been analyzing a mountain of quantitative data gathered in the Louisiana School Effectiveness Study (LSES).⁷ LSES phases three and four had included a “double blind.” The 16 schools in those phases were carefully chosen matched pairs. In each demographically matched pair, one had a multi-year history of unusually high academic performance, and the other school had a stable history of underperformance. Neither the schools nor the observers were to know which school of each pair was the “positive” or the “negative” outlier.

We eventually learned that one aspect of the design had failed: armed with no prior knowledge, 100 percent of the observers had intuited the status of 100 percent of the schools where they had observed. Stringfield read

and re-read the observers’ detailed case studies, trying to understand how the observers—most of whom were not professional educators and some of whom had not been in an elementary school since they were students—had done it. Every negative outlier school had at least one exceptional educator, and many had multiple attractive characteristics. None of the positive outliers was implementing “the latest” reform, and in fact, several looked at first blush to be as plain vanilla as schools could be.

As he ponderously distilled eight dimensions of qualitatively observed differences (eventually published as Stringfield & Teddlie, 1991), he chanced to read the then-current edition of *Smithsonian* magazine, which included a popularized article on High Reliability Organizations (Pfeiffer, 1989). HRO contained a vocabulary for explaining what was being described in the LSES case studies: in the negative outliers, a wide range of behaviors and results were tolerated, while the positive outlier schools had in common a clear focus on students’ academic achievement and intolerance for observably ineffective educator behaviors. Contrasted with the negative outliers, the positive outlier schools were much more reliable education providers.

Here are the 12 HRO principles, briefly stated and seen through the eyes of an educational effectiveness researcher:

1. Organizational reliability evolves under a particular circumstance. HROs evolve when both the larger society and the professionals involved in the

⁵In the mid-1980s, Stringfield managed Northwest Lab’s Chapter 1 Technical Assistance Center in Denver, working with Chapter 1 (Now Title I) programs across several western states. The work required regularly moving between providing professional development to teachers and paraprofessionals within and across states, working with LEAs and SEAs, and semi-annual meetings with federal officials in Washington, D.C. As would be predicted by research on loosely coupled street- (and federal-) level bureaucrats, Stringfield quickly learned “the Chapter 1 law,” as implemented, varied greatly among states, among LEAs within states, and among schools within LEAs.

⁶We believe that universities are at least partially responsible for this issue. We in universities often require few-to-no courses in learning how to differentiate between credible research and opinion, and when we do offer the courses, the practical applicability of what is offered is often limited.

⁷The quantitative and some of the qualitative results were published in a series of articles and in Teddlie and Stringfield (1993).

working of the organization come to believe that failure of the organization to achieve its key goals would be disastrous. (As noted previously, we believe that this condition is rapidly being met today.)

2. HROs require a clear and finite set of goals, shared at all organizational levels.
3. An ongoing alertness to surprises or lapses exists, and small failures in key systems are monitored closely because they can cascade into major problems. In order to sustain multi-level awareness, HROs build powerful databases. These databases possess "Four R's": relevance to core goals; rich triangulation on key dimensions; real-time availability to all organizational levels; and regular cross-checking by multiple, concerned groups.
4. The extension of formal, logical decision making analysis as far as extant knowledge allows. Regularly repeated tasks that are effective become Standard Operating Procedures (SOPs).
5. HROs actively sustain initiatives that encourage all concerned to identify flaws in SOPs and honor the flaw finders.

Because high reliability is a social construction and requires high levels of individual professional decision making, HROs perpetually engage in the following three activities:

6. Active, extensive recruiting of new staff at all levels.
7. Constant, targeted training and retraining.
8. Rigorous performance evaluation.

Four additional characteristics follow:

9. Key equipment is kept in high working order.

10. Because time is the perpetual enemy of reliability, HROs are hierarchically structured. However, during times of peak activity, whether anticipated or not, HROs display a second layer of behavior that emphasizes collegial decision making, regardless of the formal position of the decision maker.

11. Clear, regularly demonstrated valuing of the organization by its supervising and surrounding organizations. All levels work to maintain active, respectful communication geared to the key goals of the HRO.

12. Short-term efficiency takes a back seat to very high reliability.

Two additional points relate to the HRO characteristics. The first is that while these characteristics must necessarily be described separately, *their effects are presumed to be multiplicative, not merely additive*. The total absence of any one can nullify great efforts to obtain others. Standard Operating Procedures can become mindlessly rigid in the absence of ongoing honoring of flaw-finders and process/program improvers. Aggressive recruiting in the absence of supportive, long-term professional development is futile. The first 11 characteristics cannot be sustained if an organization continues a history of such poor accounting and economic prediction that it must periodically make drastic cuts in personnel, equipment, etc.

A second note concerns the description of the characteristics. It would be easy to regard each of the above HRO characteristics as existing in a stable state. In fact, all are *dynamic and regularly evolving*. As technologies advance, systems have the opportunity to create much richer databases. Last year's teacher recruiting effort, however successful, becomes the baseline for measuring this year's effort, and so on. In human

organizations, reliability is a socially constructed, evolving phenomenon.

The High Reliability Schools project is born

In 1991, Stringfield wrote a "think piece" exploring the potential for HRO principles to be used in school reform and presented it at the International Congress for School Effectiveness and School Improvement (ICSEI). David Reynolds, the founding co-editor of *School Effectiveness and School Improvement*, happened to be in the audience. A year later, Reynolds was presenting an after-school lecture to a group of British educators and briefly discussed the fact that a "mad American" had the idea that schools could be operated with the same reliability as air traffic control towers. To Reynolds' surprise, a group of educators came to him immediately after his presentation and said, "Let's do it." Reynolds asked, "Do what?" The local educators stated a desire to try to operate their schools with the remarkable reliability of air traffic controllers, and the High Reliability Schools project began.

HRS's immediate challenge was to take the abstract ideas of HROs and convert them into concrete, usable professional development segments for teachers and "heads" (in the United States, principals). As the boundaries of Stringfield's and Reynolds' skills in this area became manifest, the two contacted a colleague, Gene Schaffer, who was exceptionally skilled in translating abstract ideas regarding "effectiveness" and "school improvement" research into concrete educational professional development segments. This team has now worked together for nearly 15 years.

Three overarching sets of ideas defined the HRS project. The first was that the broadly defined fields of teacher, school, and system

effectiveness had evolved far enough to provide some level of guidance to practical educators. The second concerned our conception of the components of HROs.⁸ Third, our assumption was that the specifics were not so refined that they could be implemented lock-step in every school and classroom, but that the general principles would be able to guide local educators who were willing to work with college professors to “co-construct” (Datnow & Stringfield, 2000) a reform. We were explicit with all the local educators that we would be entering as equal partners in an exploration, or not at all. We professors would bring relevant research knowledge to the table, and the diverse local educators would work with us to make practical applications of the research.

The research knowledge bases we attempted to bring to the HRS schools were these:

1. **Teacher effects.** For example, Good and Brophy’s (1987, 2007) *Looking in Classrooms*, including training in a series of low- and high-inference teacher-effects-related observation instruments.
2. **School effects.** For example, including broad, general principles (e.g., “school climate”) and specific, alterable variables (e.g., “effective use of class time”).
3. **Extensive data gathering and use.** This was a relatively under-developed field at the time, but we encouraged schools to adopt a testing scheme that was new, and promised to provide early indications of students’ ability to perform on national tests at age 16.

4. **Popularized findings from studies of improving businesses.** Such things as adopting “big, hairy, audacious goals” (BHAGS) (Collins & Porras, 1996) instead of modest, “reasonable” goals.

We believed that the sets of findings from these fields, although valid, were being implemented around the United States and in various locations around the globe with indifferent fidelity. The Stringfield and Teddlie (1991) article gave us some reason for optimism that if the schools were willing to set very high goals, and if they were able to implement the effectiveness fields’ findings with high reliability, large achievement gains were possible. Two possible sets of tools were likely to enhance implementation reliability. The first set was the characteristics of HROs.⁹

All three members of the development/research team had participated in previous studies in which well-intended change efforts had failed to achieve strong implementations or desired outcomes. Believing that reliability would only be possible in the context of strong local buy-in, we adopted Datnow’s conception of co-construction (see also Berman & McLaughlin’s [1978] “mutual adaptation”). If teachers and school heads (principals) co-constructed the reforms, we reasoned that their ownership would be higher, as would our overall chances for success. We explicitly stated that any school that didn’t want to work hard at co-creating the reform should not participate, and that we knew a good amount about the various “effectiveness” research bases and HRO principles. However, we always paired those statements with a declaration that the world’s leading

experts on the specifics of the schools were the local teachers and heads. Success required a melding of expertise.

Groups of schools from three British LEAs asked to participate in the HRS project. We present our data from the Welsh cohort of schools, as it received the most developed version of the reform and as such, presents the most straightforward case description.

Neath-Port Talbot Local Authority: A very successful HRS project

The Neath-Port Talbot (NPT) area is located along the southern edge of Wales, with the Severn Channel as its southern boundary. NPT is about one hour’s drive due west of the Welsh capitol of Cardiff. Traditionally, the core of the area’s economy was a combination of mining and steel mills. However, the mines were closed over 20 years ago, and the one remaining steel mill is a fraction of its former self. In terms of economic deprivation (poverty), the Neath-Port Talbot area ranks 19th of 22 Welsh districts. That standing has been stable for well over a decade.

The High Reliability Schools project began in Neath-Port Talbot after Professor Reynolds made a presentation to the Welsh Secondary Heads Association. Four heads, three from NPT and one from a neighboring authority, became known as “The Old Welsh Four,” and they quickly became advocates of the project. Within a few months, all 11 secondary schools in the NPT LEA had been welcomed in, and they worked as a unified group. The intervention began in the spring of 1996 and continued for nearly four years. The Welsh agreement among the schools and the researchers was as follows:

⁸ In retrospect, it is clear that we viewed the HRO research base through the lens of teacher- and school-effectiveness researchers. There are other ways to conceptualize the HRO field (e.g., Weick & Sutcliffe, 2007), but we believe that in the context of education, the principles generally hold up.

⁹ For more detailed discussion of this conception of HRO characteristics, see Stringfield (1995); Stringfield, Reynolds, & Schaffer (2008).

- All of the schools would focus on 2–4 very ambitious goals. One required goal was a substantial 5-year rise in the percentage of students obtaining 5 or more A* to C grades on the General Certificate of Secondary Education (GCSE) tests.¹⁰ A second was improved attendance. Each school chose up to two additional goals.
- The heads (principals) would lead the efforts, and the heads and faculties would implement the HRS program schoolwide from the start of the project.
- All schools and departments within schools would agree to share successes and failures, and thus create learning communities within and across schools and LEAs. Each school and department would commit to studying “best practice,” both from the international research bases and within and without the HRS schools in England and Wales. The researchers would present school-level series of workshops on the theoretical underpinnings of “High Reliability Organizations,” and the research bases on school effects, school change, and teacher effectiveness. Armed with this knowledge, teachers would engage in within- and between-school classroom observations and “no-fault” feedback to peers. Importantly, all agreed that there would be no one piece of research or observational learning required of any school or teacher. HRS was to rely on the well-informed and supported professional judgment of practicing educators in the diverse schools.
- The researchers and administrators of each school would support the

- faculties in becoming uniquely “data rich.” Students would be given short tests as they entered the schools, and age/grade-level teams of teachers would meet and discuss how best to address each student’s needs and how to maximize each student’s chances of academic success. The resulting student-level data sets were to be (a) rich in individual students’ academic histories, (b) available to all teachers and administrators, and (c) regularly shared and discussed by all grade-level teams within schools.
- Almost all of the schools purchased a university-based system of storing and reporting initial intake and eventual GCSE scores. The system made it relatively easy for school personnel to compute a “value added” measure.
- All faculties and administrations committed to regularly review their organization and processes to create widely understood, time-saving Standard Operating Procedures, and to identify and intervene in schoolwide fashion with their pupils who appeared to be at risk of failure.
- A focus on teacher effects/peer observations began immediately. This included both professional development time to learn core aspects of the teacher effectiveness research field (e.g., Brophy & Good, 1986), and for observation in classes within and among schools.
- A strong “departmental effectiveness” component that facilitated within-school learning was emphasized as the project developed.

- Several additional components were added over time:
 - When the assessment of incoming 11-year-old students at some of the schools indicated many were entering secondary school more than two years behind in reading, an immediate effort was launched to coordinate the secondary school’s literacy programs with those of the feeder primary schools.
 - The LEA appointed a part-time “HRS Driver” to formally coordinate activities among the Welsh district’s schools. The effect was to have HRS continuously “on the radar screen” at each school and in most departments of all schools.
 - In Wales, in addition to the “broad brush” principles of HRS and the detailed organizational features of the HRS model as outlined in the components material, there was an additional focus on what came to be called “the little things that matter.” HRS meetings increasingly centered upon regular sessions in which each school explained to the whole group of Welsh schools the practical things that they had done at the “micro” level to embed the concepts and the components in the form of practical organizational features at the point of delivery of education to pupils.
 - Additional time for professional development was built into the Welsh implementation. Heads and faculties attended regionally based residential sessions (two-day meetings at a conference center) for all head teachers and HRS representatives, and also added

¹⁰ Virtually every British student sits for the GCSEs. The traditional measure of strong academic performance for a student is obtaining “5 or more A* to C” grades on the various examinations (literature, mathematics, various sciences, etc.). Although scores have risen over the last decade, under half of Welsh students obtained 5 or more A*–C grades in the mid 1990s, and those percentages have risen to over half in the first decade of the 2000s.

national residential sessions, all aimed at enhancing knowledge transfers across schools and LEAs. The Professional Development focus in Wales tilted strongly towards turning schools into “knowledge generators” rather than passive knowledge recipients. Particularly, the HRS project focused on introducing peer observation systems to permit the charting, generation, and transmission of good practice in classrooms, training some school personnel to use observation systems that were then cascaded around the entire school.

- The team’s focus upon improving schools’ capacity to be reflective about their organizational functioning and outputs was enhanced, using additional training. Examples included the provision of sessions on the statistical analysis of data and the provision of a sophisticated, relational database that teachers could access to more efficiently analyse stored grades, background information, and test scores of pupils.
- The program began to take a close interest in the effectiveness of the primary feeder schools that were generating intakes of pupils that, in the case of most schools, were regarded as unintentionally setting “low ceilings” on what it

was possible to achieve. Primary senior management teams were invited to the secondary schools’ HRS training days. One secondary school went so far as to use some of its own resources to provide a literacy coordinator to the primary schools whose students it served.

- Finally, the HRS representatives and principals received additional materials, some of which focused on topics around being effective managers of change. Additionally, bodies of knowledge to be shared with teachers were first previewed with the head teachers and HRS coordinators, allowing the leadership to be prepared to answer staff questions and ease the material into schools. Heads also selected among possible staff development alternatives based on their perceived needs of faculty and previous efforts that had been successful.

The Welsh GCSE results

The GCSEs are viewed in Great Britain as relatively high-stakes assessments, similar to Advance Placement courses in the United States. For students, a certain number of passing grades are required for such career options as becoming a policeman or postman, and a (higher) number is required for admission to various colleges. Given that all English and Welsh secondary schools have essentially open admissions (i.e.,

students are not bound to attend the school in their specific geographic area), a rising or falling standing on the percentages of students passing 5+ GCSEs can affect the number of students choosing to attend a school. In turn, teachers’ and administrators’ positions can be gained or lost. At the low end, the LEA can close secondary schools that have a persistent pattern of very low scores. Well short of that extreme, it is not uncommon for a head teacher to lose his/her job if school-level GCSE scores fall for several consecutive years.

As seen in Table 1, in the three years prior to the HRS project, 16-year-olds in the NPT LEA had scored well below the Welsh national average on the GCSEs. With one exception (discussed below) NPT’s modest standing had not raised any hue and cry at the local or national levels. NPT was a relatively deprived area, and the expectations for student performance were modest.

By the end of the intervention, NPT scores had risen essentially to the national average (48.5% vs. 49%). The gain was impressive, but being at the national average did not attract great attention. By the research team’s follow-up in 2007, NPT’s students were scoring at well above the national average (60.7% vs. 54.2%), and a nationally publicized “value added” assessment had found NPT to be by far the most “value added”

Table 1

Neath-Port Talbot Local Authority, two specific schools, and Welsh national mean percentages of 15–16-year-old students obtaining 5 or more A–C scores on the GCSEs, 1994–2007*

Time Frame/ Group	1994–1996 (pre-)	2000 (post-)	Initial Gain (pre- to 2000)	2007 (follow-up)	Longitudinal Gain (pre- to 2007)
NPT LEA	33.3%	48.5%	15.2%	60.7%	27.4%
Sandfields	14%	35%	21%	47%	33%
Cwmtawe	31%	51%	20%	75%	44%
Wales	40.7%	49%	8.3%	54.2%	13.5%

LEA in Wales (Stringfield, Reynolds, & Schaffer, 2008).

The two schools highlighted in Table 1 tell interesting longitudinal stories. Sandfields secondary is located in a very disadvantaged public housing complex. Because the school's GCSEs had been especially low for several years, the national government threatened to close it. The school head had invited Reynolds to make a presentation to the faculty to determine if there was sufficient interest in participation in the HRS project. Among the questions asked by the faculty was, "If we were to participate, where would you suggest we start?" The school facilities and grounds were in poor shape, and Reynolds suggested starting with a cleanup campaign. The faculty involved the entire community, and in a few weeks, the school's appearance was significantly improved. This gave the faculty a sense of early accomplishment, energy, and hope to go forward. Initially, the majority of the faculty would have been delighted to have achieved a 25 percent or their students obtaining 5 or more A*-C's on the GCSEs. In one year, fully 50 percent of their students earned that high standard. In 1996, the faculty would have thought achieving at that level was impossible.

The second school, Cwmtawe, is located in a more middle-class community. In the three years prior to participating in the HRS project, their level of student achievement on the GCSEs averaged 31 percent. Although this was not viewed as deeply problematic, the school's teachers—and, in particular, the administration—had higher ambitions. Probably no school embraced the HRS principles more fervently than Cwmtawe.

During the implementation years, the school raised its percentage of students obtaining 5+ A*-C's by 20 percentage points (from 31% to 51%) and for the first time exceeded the national average. By 2000, the head, deputies, and teachers had become expert at examining each student's incoming grades and test scores, and at working with the students and their families to produce multi-year plans for each student's success. The result has been that Cwmtawe's scores were well above the national average on 5+ A*-C's, and over the last decade their rate of improvement has been three times the national average. Just as impressive, the school has committed to having all students achieve passing scores on at least some of the GCSEs (in 2007, 98% of students achieved 5+ A*-C scores), and they have set a new goal of having many of their top students obtain 10+ A*-C's. So the school is focusing not just on the state-defined measure, but on high levels of success for all students set by the school. Not every school in the LEA has experienced this level of success, but it is noteworthy that 10 of the 11 secondaries in NPT produced 11-year gains that exceed the national average.

In summary, the Welsh LEA was the third to join the HRS effort and received the more nearly polished presentation throughout. The LEA provided consistent levels of support to its schools, and the heads and teachers were, on average, enthusiastic co-constructors of the reform in their community. The heads took charge of the project from the beginning and probably shared more of their frustrations and successes within and across schools than either of the other groups of schools. The results of this union of researchers—who

brought to the table valid findings from previous research—and enthusiastic, improvement-focused local educators continue to speak for themselves.

The English "leafy suburb" district: Our unsuccessful HRS pilot project

In some ways the most complex, and in several ways the least encouraging, of our British HRS LEA stories comes from the first district to suggest and implement the project. Several things went well in this LEA, but several others were problematic. Studying past failures to avoid future ones is a key process in HROs, and we discuss these as part of a prelude to discussing when and where HRS can and probably cannot help local educators improve their schools. Among the strengths in this LEA were several enthusiastic central office staff members, many fine educators in the schools, and enthusiasm inherent in being the first to try to develop a reform. Yet, the project faced several challenges that proved fatal.

The HRS program was developed "on the fly," and there were obvious rough edges on the professional development components, which certainly harmed the project's credibility.¹¹ The LEA had endorsed two separate reform efforts; eight of the 16 secondary schools in this LEA chose one reform, and the other half chose the other. In the end, neither was successful, and this lack of coordinated focus may have been a substantial problem for both reform efforts. In this initial implementation, we reasoned that the school heads already had very demanding jobs and suggested making deputy heads the "HRS drivers" of the schools. This inadvertently communicated that

¹¹ Bob Slavin, co-developer of Success for All (SFA), has observed that the first schools to implement almost all new SFA components have been among the least successful. His explanation (personal communication) has been that the rough edges reduce teacher and administrator confidence and commitment.

HRS was of secondary importance in the schools. The effort was initiated by central office staff without initial enthusiasm from individual schools. By contrast, in NPT, local heads had enthusiastically lobbied for HRS buy-in. Finally, at the time of reform implementation, the LEA was already achieving at the national average for percentages of students obtaining 5+ A*-C's, and there was no strong motivator for the schools to take on HRS, or any other demanding reform effort.

Lessons learned

Studies across a range of countries have found that producing measurable change in student achievement is more likely in elementary school reforms than secondary, yet the HRS project produced dramatic student achievement gains in secondary schools. We, as the development team, contributed a full measure to the lack of success in the initial "leafy suburban" site, but the lessons learned from it made contributions elsewhere. One clear lesson is that simply joining an HRS project isn't an automatic route to academic improvement.

Two U.S. educational improvement efforts involving HRO components

Here, we relate two efforts to use HRO principles to enhance reform efforts in the United States.

Grant County, Kentucky

Grant County is a small, rural district in North Central Kentucky, southwest of Cincinnati, Ohio, and northeast of Louisville, Kentucky. Mike Hibbert, Grant County's superintendent, heard about the British HRS project at a conference, and determined to use HRO principles to solve one of his district's more enduring problems—its high dropout rate. An analysis of their data indicated that students who dropped out were unusually likely to have repeated 9th grade.

Mr. Hibbert guided his small central office staff and leaders from Grant County's one middle school and one high school in an effort to focus on helping students succeed in their 8th to 9th grade transition, and, as a result, be more likely to succeed in 9th grade and graduate from high school. The district engaged in an extensive review of issues related to secondary school success and eventually focused on the issue of middle-to-high-school transition. They identified a substantial body of literature on necessary steps for successful transitions (Allen, Christian, & Hibbert, 2010; Morgan & Hertzog, 2001; Oakes, 2009).

A large team of middle- and high-school teachers and administrators took a range of steps, including enhanced student and parent involvement, teacher "interventions" between the schools, and 8th-grade student days spent at the high school, all carried out using HRO principles and processes. The results were immediate and dramatic. Hibbert (2010, personal communication) reported that whereas freshman retention had been 38 percent pre-intervention, in the 2009–2010 school year, "the retention rate was 1 percent."

Kentucky mandates and funds all high school students taking the ACT's "PLAN" (pre-ACT) test in 10th grade. In the second year of transition implementation (when the first implementation cohort reached 10th grade and took the PLAN), Grant county's PLAN scores rose a full point (equivalent to a rise of 50 points on the SAT). Further, Hibbert reports that the number of failing grades are down significantly among 9th grade students at his high school.

The Effective Schools for the 21st Century (ES-21) Project

In 2004, the Olin Foundation funded an effort to conduct a "gold

standard" (random assignment) study of the implementation of "Effective Schools" variables and processes (Taylor & Bullard, 1995; Teddlie & Reynolds, 2000) in a sufficient number of schools across several states so as to demonstrate the continuing validity of those principles at the standards required by the federal "What Works Clearinghouse." The implementation team decided to use HRS principles as over-arching principles of implementation in the project. The project confronted a continuing series of complications, including these:

- Many LEAs were interested in participating, but attempts to ensure random assignment resulted in over a dozen districts declining offers to participate, even though the project was 100 percent externally funded.
- The requirement of random assignment after they agreed to participate left many principals and teachers feeling that the project was more someone else's research than their own reform.
- No district had as many as 50 percent of its elementary schools participating in the "experimental" group. In several instances, central office staff "borrowed" key ES-21 principles and presented and/or implemented them in control sites.
- Without all schools from a district participating in the experimental component, several superintendents became less than enthusiastic about the idea of some of their schools getting something that others were not.
- The experimental LEAs and schools experienced high rates of professional staff instability, such that re-training became a norm in the project. Four of five LEAs experienced at least one

superintendent turnover; the 17 experimental schools had a total of 35 principals over the three years; some school leadership teams experienced between 100 percent and 200 percent turnover.

Qualitative follow-up interviews consistently found teacher and principal enthusiasm for the project, with educators regularly reporting that they were better prepared to deal with future changes as a result of ES-21 participation. However, in the end, the project did not produce achievement test score gains for the two carefully followed cohorts of students in the study.

Discussion

It is hardly surprising that most school reforms fail. Complex systems—such as schools and school systems—that are inadequately understood and modeled are unlikely to be successful hosts for reforms of almost any type. We assert that research on very complex systems that must succeed in their core missions the first time every time (e.g., High Reliability Organizations) offers guidance for school reform. Efforts to use HRO principles to guide reform in several contexts offer both hope and cautions for future educational reformers:

1. Dramatic improvements in student outcomes are possible, and possible at scale.
2. To achieve those results, the reform components must themselves be based on substantial bodies of research that have demonstrated their value in improving student performance.
3. In educational reform, as in research, “reliability sets the upper boundary of measured validity.” HRO research can play a critical role in producing Highly Reliable Schools.

4. Among the conditions necessary for HRSs to evolve are the following:

- As a first condition, both the public and the professional educators must realize that in the 21st century, the costs of educational failure are catastrophic for the individual students who do not achieve their full potential and for the rest of us in society. This is a dramatic shift from 50 years ago.
- Buy-in from both the LEAs and the schools’ leadership to a focused set of goals is a critical next step.
- There must be a perception existing, or created early on, that failure to achieve core goals is unacceptable.
- An understanding of—and openness to—the idea that any reform, including HRO-based reform, is a combination of external ideas and continuously evolving local contexts. Just as there are no two air traffic control towers that are alike, there are no two schools needing exactly the same reforms, the same Standard Operating Procedures, and so on. Further, any one school’s need for any one SOP may change over time. Dynamic organizations must be dealt with dynamically.
- A minimal level of leadership stability, combined with carefully targeted leadership transitions, is necessary to sustain reliability.
- The HRO characteristics described earlier in this chapter must be followed in detail.

Viewed from an HRS perspective, the conditions that predict reform failure also are knowable. They include, but are not limited to, the following:

1. A lack of initial buy-in to the idea that dramatic improvement in student outcomes is possible.
2. Too many diverse goals.
3. Attempting to implement reforms that are not clearly informed by rigorous research.
4. Lack of multi-year commitment to intensive, shared professional development.
5. Leadership and staff instability, especially if not accompanied with careful, real-time induction into HRS principles.

Conclusions

Complex systems that are inadequately understood and modeled, such as schools and school systems, are unlikely to be successful hosts for reforms of almost any type. It is hardly surprising that most school reforms fail. In this chapter, we have argued that research on very complex systems that must succeed in their core missions the first time, every time (High Reliability Organizations) offers guidance for school reform.

Our first overarching conclusion is that the conditions now exist in which substantially higher educational reliability in the United States is possible. The costs of failure—both for the individual and the society—have become too great for unreliability to continue. Hence, we believe that the country’s fundamental choice is not whether to become more reliable, but whether to stumble forward, feeling our way and making many, many mistakes; or whether to understand and control a more efficient process of increasing educational reliability.

The second conclusion allows for a good deal of optimism. Our data indicate that, under specifiable conditions, High Reliability Organization principles can be productively applied in school and district contexts. Thoughtfully, consistently applying HRS principles has produced dramatic results in the United Kingdom and the United States, and could do so again in other schools and districts.

HRS models may initially appear to have a “mechanistic” feel to them. Nothing could be further from the truth. In many schools and indeed countries, schools and systems are “tight” on the processes that are meant to exist and “loose” on the systems to achieve these processes. HRS is the opposite—loose on the precise organizational processes needed, leaving those to be determined in detail by local educational professionals. Where HRS is tight is on specifying the concepts and systems that schools should use to generate their often-different processes.

In an educational world where school systems too often tell their teachers what to do, the HRS model is representative of a different philosophy which sets schools free to determine which research- and proven-practice-based practices to implement. This is surely the way to create a more informed, effective, better supported, and more reliably successful educational profession.

References

Allen, D., Christian, C., & Hibbert, M. (2010). *Lost in transition: A grade 9 transition program using articulation activities*. Louisville, KY: University of Louisville.

Berman, P., & McLaughlin, M. (1978). *Federal programs supporting educational change, Vol. VIII: Implementing and sustaining innovations*. Santa Monica, CA: Rand.

Brophy, J. E., & Good, T. (1986). Teacher behavior and student achievement. In M. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed.) (pp. 328–376). New York, NY: Macmillan.

Collins, J., & Porras, J. (1996.) Building your company’s vision. *Harvard Business Review*, 74(5), 65–77.

Datnow, A., Lasky, S., Stringfield, S., & Teddlie, C. (2006). *Systemic integration for effective reform in racially and linguistically diverse contexts*. New York, NY: Cambridge University Press.

Datnow, A., & Stringfield, S. (2000). Working together for reliable school reform. *Journal of Education for Students Placed At Risk*, 5(1&2), 183–204.

Good, T., & Brophy, J. (1987). *Looking in classrooms* (4th ed.). New York, NY: Harper & Row.

Good, T., & Brophy, J. (2007). *Looking in classrooms* (10th ed.). New York, NY: Harper & Row.

Goodlad, J. (1984). *A place called school*. New York, NY: McGraw-Hill.

Heckman, J. J., & LaFontaine, P. A. (2010). The American high school graduation rate: Trends and levels. *The Review of Economics and Statistics*, 92(2), 244–262.

Hibbert, M. (2010, October 24). Transition data for Grant County schools. Personal communication.

Land, D., & Stringfield, S. (2005). Educational governance reforms: The uncertain role of local school boards in the United States. In N. Bascia, A. Cumming, A. Datnow, K. Leithwood, & D. Livingstone (Eds.), *International handbook of educational policy* (pp. 260–280). New York, NY: Kluwer.

Lipsky, M. (1980). *Street-Level Bureaucracy*. New York, NY: Russell Sage Foundation.

Morgan, L., & Hertzog, C. (2001). Designing comprehensive transition plans. *Principal Leadership*, 1(7), 10–18.

No Child Left Behind (2001). Public Law 107-110. Retrieved from <http://www2.ed.gov/policy/elsec/leg/esea02/107-110.pdf>

Nunnery, J. A. (1998). Reform ideology and the locus of development problem in educational restructuring. *Education and Urban Society*, 30(3), 277–295.

Oakes, A., & Waite, W. (2009, May). Middle-to-high school transition: Practical strategies to consider. *The Center for Comprehensive School Reform and Improvement Newsletter*. Retrieved from http://www.centerforcsri.org/index.php?option=com_content&task=view&id=669&Itemid=5

Pfeiffer, J. (1989). The secret of life at the limits: Cogs become big wheels. *Smithsonian*, 20 (4), 38–48.

Reynolds, D., Creemers, B., Stringfield, S., Teddlie, C., & Schaffer, E. (2002). *World class schools: International perspectives on school effectiveness*. New York, NY: Routledge/Falmer.

Stringfield, S. (1991, January). *A model of elementary schooling effects: Evidence of support*. Paper presented at the meeting of the International Congress for School Effectiveness and Improvement, Cardiff, Wales.

Stringfield, S. (1995). Attempts to enhance students’ learning: A search for valid programs and highly reliable implementation techniques. *School Effectiveness and School Improvement*, 6(1), 67–96.

Stringfield, S. (2008). School boards and raising student outcomes: Reflections (Confessions?) of a former urban school board member. In T. Alsbury (Ed.). *The future of school board governance:*

Relevancy and revelation. Lanham, MD: R & L Education.

Stringfield, S., Reynolds, D., & Schaffer, E. (2008). Improving secondary students' academic achievement through a focus on reform reliability: 4- and 9-year findings from the High Reliability Schools project. *School Effectiveness and School Improvement*, 19(4), 409-428.

Stringfield, S., Schaffer, E., Lasky, S., Overman, L., Chrispeels, J., Burke, P., Daly, A., Bolivar, J., Murga, S., & Johnson, P. (2010). *The four quartets: Effective Schools for the 21st Century Project final report*. Louisville, KY: University of Louisville.

Stringfield, S., & Teddlie, C. (1991). Observers as predictors of

schools' multi-year outlier status. *Elementary School Journal*, 91(4), 357-376.

Stringfield, S., & Yakimowski, M. (2005). The promise, progress, problems, and paradoxes of three phases of accountability: A longitudinal case study of the Baltimore City Public Schools. *American Educational Research Journal*, 42(1) 43-76.

Taylor, B. (2002, January). The effective schools process: Alive and well. *Phi Delta Kappan*, 375-378.

Taylor, B., & Bullard, P. (1995). *The revolution revisited: Effective schools and systemic reform*. Bloomington, IN: Phi Delta Kappa.

Teddlie, C., & Reynolds, D. (Eds.) (2000). *The international handbook of*

school effectiveness research. London, UK: Falmer.

Teddlie, C., & Stringfield, S. (1993). *Schools make a difference*. New York, NY: Teachers College Press.

Weick, K. (1976). Educational organizations as loosely coupled systems. *Administrative Science Quarterly*, 21(1), 1-19.

Weick, K., & Sutcliffe, K. (2007). *Managing the unexpected* (2nd ed.). San Francisco, CA: Jossey-Bass.

To ask questions, or for more information, write to the authors at their e-mail addresses:

Sam Stringfield: sam.stringfield@louisville.edu

David Reynolds: david@

davidreynoldsconsulting.com

Eugene Schaffer: schaffer@umbc.edu

Chapter Three

High Reliability & Leadership for Educational Change

By G. Thomas Bellamy, University of Washington Bothell

Introduction

On April 20, 2010, a deepwater drilling platform used by BP exploded, releasing an 86-day torrent of oil into the Gulf of Mexico and wreaking havoc on the natural environment and on the lives of millions of people, their families, communities, and economies. The aftermath will be felt for generations. When consequences are this catastrophic, the public expects organizations to work without failure, and headlines follow when they do not.

A school that fails may lack the visibility of a BP disaster, but it can have its own catastrophic human consequences for the lives of young people, their communities, and our society. The public understands this all too well, and its high expectations for reliable school performance are reflected in government policymaking and in local news. In these respects, students failing in school might not be all that different from a petroleum company's off-shore accident. As such, the BP disaster might have at least one positive outcome: providing public education with insights about how to avoid failure.

Thanks to research on many accident-free organizations, we know a great deal about managing for success in uncertain situations. These High

Reliability Organizations include familiar enterprises such as air traffic control, chemical manufacturing, air travel, electric power generation, and wildland firefighting groups. All face serious hazards in unpredictable circumstances.

HROs achieve reliability through four distinctive organizational accomplishments:¹

1. Sustaining commitment to a dual bottom line
2. Centralized procedural control and standardization
3. Flexibility for situational improvisation
4. Combining opposite operating modes

Assuming that all structures and strategies are fallible, leaders of successful HROs build the capacity to work in two modes, one standardized and centrally controlled, the other decentralized, improvisational, and situation specific. Each way of operating is always ready to use as shifting circumstances either allow normal work toward organizational goals or threaten reliable performance.

As our understanding of HROs expands, so too does interest in applying their strengths and

strategies to manage uncertainty and improve performance in many other organizations, including education (Bellamy, Crawford, Huber-Marshall, & Coulter, 2005; Weick & Sutcliffe, 2007). Stringfield and his colleagues (Stringfield & Datnow, 2002; Stringfield, Reynolds, & Schaffer, 2010; Stringfield, & Yakimowski-Srebnick, 2005) already have shown that HRO strategies, used in combination with concepts from the effective schools literature, can support dramatic and sustained improvements in school learning outcomes.

Of course, public schools are quite different from most HROs. Educators work in public organizations that are naturally open to outside influences; their work is people-and relationship-intensive and depends on far less prescriptive knowledge. Thus, while high reliability seems important to leadership for educational change, contextual differences mitigate against uncritical transfer of HRO practices to schools.

1. Sustaining commitment to a dual bottom line
- Successful HROs find ways to balance simultaneous commitments to achieving desired results while avoiding accidents or failure. They are adept at "finding a balance between

¹ Similar to the research on effective schools, qualitative studies of HROs have led to several taxonomies of the distinctive characteristics of these organizations, each highlighting slightly different aspects of reliable operations (e.g., see Roberts, 1990; Rochlin, 1993; Weick & Sutcliffe, 2007). The categories used here highlight challenges in applying HRO strategies in schools.

today's profits and tomorrow's potential disasters" (Roberts, Bea, & Bartels, 2001); they "consider reliability as important an outcome as productivity" (Roberts & Libuster, 1993, p. 16); and they are able to "restate goals in the form of mistakes that must never occur" (Weick & Sutcliffe, 2007, p. 151). They are equally committed to providing service and avoiding failure, based on strong agreement about both the definition and value of success and costs of failure (LaPorte, 1996). For HROs, such dual priorities are not just slogans. The underlying values—commitments to what the organization needs to accomplish and what it should never allow to happen—become cornerstones for the organization's culture (Weick, 1987).

News coverage² of BP's accident provides some insight into just how difficult it can be to sustain a cultural commitment to this dual bottom line. Confronted with its history of several prior accidents and safety violations, BP's then-CEO began his tenure with the promise that "the company would make safety its number 1 priority" (NYT, 2010, July 12)," and maintained that he had been "laser-focused" (NYT, 2010, June 18) on safety while leading BP. But reports of priorities on the ground sound quite different: "Taking shortcuts was ingrained in the company's culture, and everyone in the oil business knew it" (NYT, 2010, June 18); "BP was developing a reputation as an oil company that took safety risks to save money" (NYT, 2010, May 31).

Why such a discrepancy between management's stated priorities and the operating culture? One possibility is the seemingly disconnected communication about goals for productivity and safety. News coverage can be incomplete, of

course, but one set of messages seemed to place an absolute priority on safety, while another attested to the company's total commitment to productivity and profitability ("BP tries to Reassure Shareholders," NYT, 2010, July 7). There is little to no evidence of shared company understandings about a balanced commitment to safety and productivity and what this balance means for organizational routines and relationships.

Implications for leading school change: "Balanced and sustainable goals"

Balanced goals are just as important for schools as for high-risk industries. Pressure for educational excellence comes from families and local communities, state and federal policies, and the commitments of educators themselves. The desire to benchmark outcomes against the best in the world (Barber & Mourshed, 2009) simply reinforces these aspirations. At the same time, the language of educational improvement reflects pressure to avoid all academic failure, with its emphasis on eliminating achievement gaps, preventing dropouts, setting non-negotiable goals, and so on. Just like off-shore oil drilling, air traffic control, or hospital operations, it is insufficient for schools to point to the success of some students, however impressive, while others experience failure. *High reliability learning*—bringing all students to proficiency regardless of their circumstances and our challenges—has joined high academic achievement as a paired expectation for public school success. In fact, the escalating requirements for adequate yearly progress increasingly make high reliability learning a precondition for schools to provide other educational services, just as safe operation is a prerequisite

for continued operation of most HROs. Like HROs, schools achieve the benefits of balanced goals only when these espoused goals are supported in the school's underlying cultural values. Establishing and sustaining goals in public education is a complex process that requires ongoing engagement by those leading change.

State and federal policies offer strong incentives to define school goals in terms of standardized-test scores, but families and communities expect much more. "We want it all" was John Goodlad's (1984) summary of extensive national research on expectations of schools. In addition to academic learning, communities count on public schools to ensure students' safety and well-being; support social, civic, and ethical development; and to help students pursue individual talents and interests. Since family circumstances and children's needs vary, different priorities emerge from the many demands competing for a school's limited time and resources. And, as public institutions, schools cannot simply settle these priority conflicts through administrative fiat. Instead, the priorities for what schools should achieve and avoid are decided through continuing dialogue, in both the internal and external communities of the school district, which is punctuated by school board elections, funding ballots, labor negotiations, and leadership changes.

In this context, educational leaders face contradictory requirements in their efforts to establish and sustain balanced goals. School goals for what to achieve and avoid must be open to change as a result of ongoing, honest dialogue and political decision making. But a school's goals also must be stable enough to provide a

² All news quotations are from the *New York Times* and are available as a set at http://topics.nytimes.com/top/reference/timestopics/subjects/o/oil_spills/gulf_of_mexico_2010/index.html.

foundation for the learning goals, performance indicators, and student assessments that guide the details of change management (Marzano & Waters, 2009). The capabilities needed to pursue any set of achievement and avoidance goals are incorporated over time into the organization's structures, staff skills, and culture. The leader's task is to open opportunities for participation and create the framework for productive discussion (Chrislip, 2002).

Of course, even when goals are developed through broad participation and deliberation, school leaders have the challenge of fostering internal coherence. This is challenging because of the sheer number and variety of educational goals and the opportunity costs associated with any particular set of priorities. Credible communication about balanced goals depends on open discussion of the hard questions about tradeoffs when goals for achievement and avoidance conflict.

For example, when a child is having difficulty with an arithmetic concept and needs extra time, where will that time come from? From the science or art lesson? When the teacher spends extra time with the struggling student, do other children who have mastered the concept miss out on whatever accelerated opportunities might otherwise have been available? Does everyone implicitly agree that there are some activities that cannot be displaced by extra instruction in core subjects?

School leaders may be tempted to avoid the issues, or to leave it up to teachers to decide on a case-by-case basis, but lack of clarity about priorities can also mean lack of timely action to respond to early warnings of failure. Credible communication about real priorities and tradeoffs can also help build the will and capacity

to act. It is one thing to have general agreement with a set of priorities and another entirely to reach shared commitment to resilience and a belief that staff and leadership will do everything possible to meet the school's goals for achievement and avoidance.

In their application of HRO strategies to educational change, Stringfield, Reynolds, and Schaffer (2010) asked schools to focus on a very small number of critical goals. Here we suggest two additions: (a) that school goals should explicitly address what should never happen as well as what should be achieved, and (b) that leadership for change should include ongoing commitment to conversation and stewardship of the goals, both internally and externally, in order to preserve balance and achieve stability in school priorities.

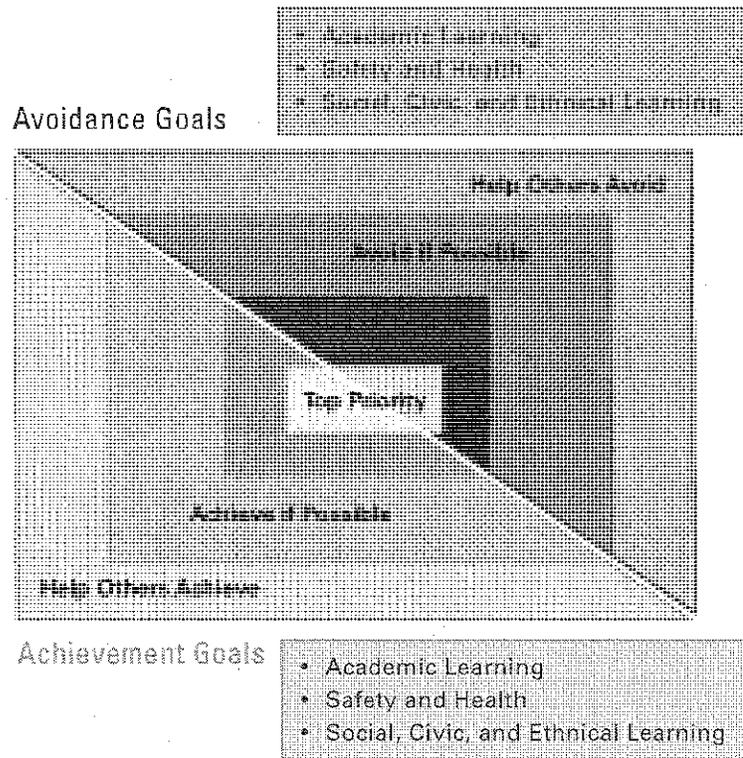
Figure 1 depicts these two additional strategies. The same framework might serve as a way to structure ongoing feedback about how well the school is perceived as implementing those priorities.

2. Centralized procedural control and standardization

Much of the literature on HROs describes organizations with centrally controlled and highly standardized operating procedures. Ways of doing things are designed by experts who can translate the field's best knowledge into practice and supported by management practices, hierarchical decision making, regular feedback, and employee incentives for following those practices. Standard procedures allow the field's current knowledge and the organization's prior learning to be imbedded in routines, provide

Figure 1

Visual prompt to support deliberation about school priorities



a way to coordinate various parts of a complex organization's work, clarify decision authorities and premises, and create a foundation for continuous improvement as processes are evaluated and refined. Successful HROs take advantage of standard operating procedures and central control by continually improving the procedures themselves and the organization's implementation capacity through training, feedback systems, and post-action reviews (Rochlin, 1993).

Although standardization and central control are useful, research on HROs makes it clear that standard procedures are insufficient to achieve high reliability. In HROs, standard procedures are implemented with an assumption of fallibility, constant attentiveness to what might go wrong, and simultaneous investment in the capacity to respond differently when the inevitable problems arise.

Press coverage of the BP accident highlights two challenges related to procedural standardization. The first is whether BP's procedures actually incorporated the best current knowledge or were designed instead to achieve efficiencies by taking greater risks. For example, one report quotes from Congressional correspondence that "some of the decisions appeared to violate industry guidelines and were made despite warnings from BP's own employees" (NYT, 2010, June 14). The second challenge is evident in reports that interpreted the failure to standardize procedures as evidence that BP was not learning from previous mistakes. An earlier BP project was described as having "cramped, chaotic conditions..." "It was like having the plumbers, the electricians and the bricklayers come to a construction site at the same time as they are laying the concrete. This was not methodical" (NYT, 2010, July 12).

Yet, despite a series of accidents and safety warnings, BP continued to rely on a case-by-case approach, rather than adopt standard operating procedures. While the high reliability literature is more skeptical than the authors of this particular news report about standard procedures as the only strategy to avoid accidents, much of the literature suggests that in most circumstances, standardization is a prerequisite for reliability.

Implications for leading school change: "Skeptical standardization"

The high level of standardization typical of most HROs is seldom present in schools, but the experience of HROs could be more relevant than first appearances suggest. Pressures to achieve high reliability learning are leading an increasing number of schools and districts toward greater central control and standardization in curriculum and instruction. In practice, school leaders standardize normal operations by adopting curriculum materials, pacing guides, common assessments, and so on, which establish consistent instructional practices across a school or district. They also build capacity to implement these standard procedures through staff selection, training, coaching, and evaluation. In such approaches, teachers are expected to comply with and build skills for established programs, communicate with supervisors about problems, and use data to improve program implementation.

Literature on HROs offers two main reasons for procedural standardization. Both seem relevant to public education but application involves special challenges. The first reason is to ensure that the field's best knowledge is applied in any given classroom. Robinson, Hoepa, and Lloyd (2009) call such procedures "smart tools," emphasizing that research knowledge can be built into

a set of procedures and routines such that not every user needs to know all of the underlying theory and data in order to receive the benefits. But every teacher knows what an honest critique of research shows: none of our smart tools—no curriculum, program, textbook, or instructional method—works reliably with all students. Thus, there is every reason to argue that standardization alone will not create fail-safe schools. Further, our field's knowledge base contains many different, often contradictory, methods for achieving educational goals (Donmoyer, 1996). And, possibly because different approaches work best with different groups of students, schools are often characterized by intense personal and professional commitments to different curricula and teaching strategies. Conflicts over which program or approach to select are practically assured whenever schools attempt standardization, and such conflict can easily derail any efforts to implement a standardized program. Critics of standardization of instructional procedures also point out that it often causes disruptive shifts during central leadership transitions and is ineffective in supporting teacher learning and development (Levine & Marcus, 2007). Given these limitations, it is reasonable to ask if standardization really makes sense in public education.

This question leads to a second reason for standardization in HROs. From extensive experience consulting with hospitals to reduce medical errors, Resar (2006) observes that, even when more than one research-based approach is available for a given clinical procedure, medical errors increase when each physician continues to use her or his preferred strategy. It's not that one is necessarily better than the others, but rather that, without reasonable standardization, the organization cannot build the

most effective systems—the staff skills, equipment, supplies, scheduling, and other components—that support the clinical procedure. So even when more than one procedure is supported in research, Resar argues that high reliability involves picking one, then building the capacity to use it well.

The circumstances that Resar (2006) describes in hospital care appear analogous to education, where some teacher discretion to choose among research-supported procedures is often expected. Resar's logic suggests, then, that skeptical standardization can be useful as one part of a leadership strategy for school change, as long as efforts to standardize pay special attention to three issues.

- First, standardization makes the most sense when the chosen procedure or program is expected to succeed with a significant majority of students and can be implemented with sufficient slack so that teachers have time to pay special attention to the inevitable group of students for whom the procedure was not successful.
- Second, standardization makes sense when the selected procedures will be used as the foundation for building capabilities in instructional materials, technology, professional development, data systems, and other resources that support instruction.
- Third, precisely because of the limitations of any particular curriculum or procedure, standardization makes the most sense in schools when it is combined with the ability of HROs to shift quickly to in-school improvisation as soon as difficulties arise.

In short, skeptical standardization is most useful when accompanied by its opposite, flexibility for teacher decision making.

3. Flexibility for situational improvisation

A third major accomplishment of HROs is their ability to operate in a flexible, decentralized, and improvisational mode when the need arises. In this mode, critical decisions about what to do are made where the work occurs, not by distant engineers or managers. When seeking expertise in understanding and responding to emerging situations, communication with colleagues replaces vertical reporting (LaPorte, 1996). Knowledge of the situation, combined with expertise to interpret situational developments, replaces general knowledge, standard procedures, and hierarchical authority as the guide to action (Roberts, Yu, & van Stralen, 2004). Operating this way requires open communication, so that those with needed expertise have an opportunity to hear about situations as they develop (Roth, Multer, & Raslear, 2006). To build capabilities for flexible operations, HROs deliberately sustain diverse perspectives and expertise on their staffs, create opportunities for employees to expand professional boundaries, and support norms of resilience that motivate responses to all threats of failure (Weick & Sutcliffe, 2007).

Such flexibility is not freelancing. Responding to situational changes in HROs is a collective process that occurs within constraints of organizational values, collaboration, and previously established decision-making routines (Bigley & Roberts, 2001). The literature contains

multiple examples of how otherwise highly controlled organizations also operate in this more flexible mode when needed, including shifts to decentralized and on-site decision making in organizations such as automobile manufacturers (Alder, Goldoftas, & Levine, 1999), aircraft carriers (Roberts, Yu, & van Stralen, 2004), and skyscraper construction (Gawande, 2009).

Implications for leading school change: "Constrained improvisation"

If skeptical standardization enables teachers to use the *profession's accumulated knowledge* in their work, then constrained improvisation³ enables them to take advantage of *contextual knowledge*—understanding of particular students, families, social groups, and evolving situations that affect learning in a specific classroom and school. In practice, operating in this mode begins with shared goals for student learning and gives teachers the authority to adapt classroom procedures as needed to achieve those goals. School leaders respect the natural differences in instructional approaches that result as teachers respond to current circumstances and implement non-prescriptive strategies for professional development and teacher evaluation.

At first glance, flexibility for situational improvisation seems familiar in schools. After all, behind their classroom doors teachers have traditionally been able to operate as they believed best, while administrators were expected to buffer instructional practice from external pressures. As suggested above, however, HROs use this operating mode in very specific ways that constrain individual flexibility. Improvisation is largely collective work, guided by clear and shared

³ Originally used in the HRO literature by Bigley and Roberts (2001), the term "constrained improvisation" seems particularly suited to the context of work in public schools, where so many different groups are empowered to influence instructional practice.

commitments to reliable performance, accountability among colleagues, and supported by appropriate checks and balances. In fact, too much workplace discretion in the absence of constraining policies and checks is associated with higher risk of failure (Roberts & Libuster, 1993). Improvisation requires no less coordination and accountability than standardized procedures; it simply achieves these results in different ways.

In schools, collective innovation depends on regular interactions with colleagues, not closed classroom doors. It is based on common goals and involves shared accountability for results. And, like other forms of teacher collaboration, collective innovation often requires administrators to develop new organizational routines that give teachers the time and support to build professional connections with a diverse group of colleagues.

The HRO literature adds an important element to existing literature on teacher collaboration. Much current commentary and procedural recommendations focus on collaboration as a means of teacher learning as a strategy for ongoing improvements in instruction (Levine & Marcus, 2007; Lieberman & Miller, 2008). Constrained improvisation in HROs begins with the more immediate concern of how to respond to a specific situation that threatens reliable student learning. Guided by a commitment to resilience, such collaboration helps to make sense of an immediate situation, develop alternative responses, and make just-in-time adjustments until the threat is addressed. Collaboration for immediate resilient action is not incompatible with longer term teacher learning, nor is it assured when teacher learning is the primary focus.

With these challenges in mind, using constrained improvisation as one part

of a strategy for leading school change involves establishing organizational constraints that make that flexibility both collective and accountable while building the capacity to work in a flexible and improvisational mode. To establish the constraints that guide this kind of flexibility, school leaders can do the following:

- Foster shared assumptions and commitments to the school's goals, reliability expectations, commitments to resilience, and considerations for decision making. These elements of school culture can help to coordinate and focus improvisational work without having to rely on a centralized authority structure (Weick, 1987).
- Create school routines and work groups that involve teachers in frequent face-to-face communication about learning challenges. This offers a context for conversations about emerging problems and allows members of a group to bring diverse viewpoints to a collective consideration of planned actions (Gawande, 2009).
- Enhance accountability for improvisational work through post-action reviews, when colleagues can reflect on what was learned from working through difficult situations (LaPorte & Consolini, 1991).

To build the school's capacity to operate in this flexible and improvisational mode, principals can apply three major HRO strategies.

1. Sustain a variety of perspectives and encourage expression of diverse viewpoints.

Schools too often overlook or deliberately discourage diverse approaches in a push to achieve a commitment to an already selected curriculum or instructional approach. Efforts to "get everyone on the

same page" may well implement a chosen program, but they will likely weaken the school's ability to operate effectively in an improvisational mode. An immediate challenge for many principals is to be visibly committed to sustaining the knowledge and commitments of those teachers who prefer and are skilled in approaches that are different from the school's adopted programs. It can be tempting to frame these differences as performance problems and pursue transfers or personnel actions. But, these same individuals can contribute important perspectives to the school's improvisational efforts to address inevitable failures in the standard programs.

2. Foster the development of teachers' informal networks and focus communication within these networks on issues of teaching and learning.

Possible approaches include professional learning communities (Dufour & Eaker, 2005), informal networks (Bidwell, 2001), teacher leadership (York-Barr & Duke, 2004), and organizational routines that require regular collective discussion of academic work (Spillane, Mesler, Croegaert, & Sherer, 2009).

3. Establish structures and routines that support collective improvisation whenever needed.

For example, grade-level teams or secondary departments that normally operate to support implementation of standardized programs might also be charged with the responsibility to provide collegial support whenever needed to respond to a student's emerging learning difficulties. Similarly, roles for instructional coaches or district-based curriculum specialists could be designed with dual responsibilities for supporting implementation of standard programs and helping teachers mobilize an early response to impending student failure.

4. Combining opposite operating modes

The two operating modes of HROs are contradictory in many respects. They depend on different organizational structures and routines, contrasting approaches to staffing and training, and different sources of authority for operational decisions. A closer look, however, shows that these two modes also depend on each other. With the pace and complexities of most modern workplaces, employees simply lack the time to devise unique solutions for every circumstance. Some standardization is necessary to provide the slack needed to respond creatively to the most difficult situations. On the other hand, as knowledge becomes more complex, few procedures work in all situations, and even when it seems possible to anticipate every contingency, the result is a system so complex that it can lead to implementation errors that require situational improvisation (Perrow, 1967; 1984). Not surprisingly, the ability to combine opposite operational modes like standardization and flexibility is closely linked to organizational effectiveness (Cameron, Quinn, DeGraff, & Thakor, 2006).

HROs are distinctive in the way they combine these contrasting operational modes. They continuously develop the capacity to operate in both modes and shift between the two approaches as situations arise to threaten reliable performance. HROs use standardized procedures for most normal operations associated with achieving expected results (goals for achievement). Then, operations shift to give situational flexibility when the organization's reliable performance is threatened (goals for avoidance), so that situational sense-making and decisions about action are made where the threatening situation exists. HROs accomplish this shift from

standard to flexible operating modes by constantly searching for procedural flaws and situations in which standard routines are unlikely to work. HROs create conditions in which employees notice and communicate about early-stage problems that could threaten reliability and establish incentives for reporting difficulties, even when they might have been caused by the person reporting the difficulty (Roberts, Yu, & van Stralen, 2004). They foster communication channels that allow information about emerging risks to be shared quickly and widely, have shared understandings about when a shift away from standard operating procedures is appropriate, and have ready-to-use structures and routines that help to coordinate the work when shifts are made to the more flexible operating mode (Bigley & Roberts, 2001). Individual employees support this rapid identification by attending carefully to operational details, describing anomalies within an informal network of peers, listening in on others' concerns, and soliciting alternate viewpoints when issues arise (Barton & Sutcliffe, 2009).

Returning to BP's response to the oil spill, news coverage of evolving events highlight the difficulties associated with such shifts between operating modes. The reports prompt one to wonder first about the skepticism that builds attentiveness to emerging problems ("Nobody believed there was going to be a safety issue...", NYT, 2010, May 29). Then, as problems became apparent, one asks what would have been required to stop the momentum of daily work toward deadlines long enough to consider alternatives. BP's employees apparently noticed many problems as events cascaded toward disaster ("Documents Show Early Worries about Safety of Rig," NYT headline, 2010, May 29; "BP Ignored the Omens of Disaster," NYT headline, 2010, June 18). But news

reports suggest that this information was never interpreted in ways that prompted a shift to on-site authority for problem solving.

Implications for leading school change: "Public warnings and orderly transitions"

Several ways of combining skeptical standardization and constrained improvisation already exist in public education. In some districts, the boundary between central control and local flexibility is simply the ragged edge of cumulative labor negotiations about whether administrators or teachers should control various decisions. In others, the remnants of organizing schools as loosely coupled systems are evident in the use of central control in some visible aspects of schooling while protecting flexibility in internal classroom operations. More recently, one frequently finds tightly standardized strategies applied to subjects that are included in a state's accountability system combined with more flexibility for teachers to exercise discretion in other subjects (Spillane et al., 2009). As an alternative to using either one approach or the other for various functions, HROs offer the possibility that a school could take advantage of the strengths of both approaches in all of its operations. As with other characteristics of HROs, however, application in education presents unique challenges.

Public warnings

Schools are rich with information about early-stage learning problems, even without waiting for the results to show up in formal data systems. Most teachers already know long before formal testing which students are advancing too slowly to meet expectations, or are being held back by the pace of instruction. But schools are seldom much better than BP at sharing this information, making collective sense

of it, and empowering teachers to work together on solutions. The first leadership challenge, then, is to overcome norms of privacy and autonomy that can limit this communication about emerging learning problems. For example, a school might agree on a shared commitment that no child would fall more than one or two weeks behind peers without collegial discussion of alternatives, so that schoolwide expertise could be tapped quickly in a search for alternative procedures.

A second challenge to effective public warnings results from the structure of specialized programs and professional roles, which can lead school staff to label emerging problems as characteristics of children, rather than results of school procedures. As teachers make sense of learning difficulties, it is easy to jump to the categories for which funding exists, and then hope that specialists can solve the student's problem. Naturally, this chain of responses can limit the range of perspectives about an emerging situation and reduce on-the-spot experimentation and adaptation.

Perhaps the most difficult challenge to an effective early warning system in schools is the *frequency* with which emerging learning problems are encountered. In much of the HRO literature, problems that threaten the organization's reliability are infrequent events. Unexpected fires break out or escape their boundaries, or rare equipment failures require sudden changes in flight plans, and organizations respond with episodic shifts in operating modes.

In schools, student learning difficulties require no less creativity in responding, but these problems occur daily. It is a rare lesson that engages all students and helps each to develop the intended knowledge and skill. In this context, it is unreasonable to assume that more than a small

fraction of early warnings will be shared collegially in the course of normal informal conversation during the school day. Consistently sharing information about many emerging problems requires a more systematic approach. A possibility is suggested in Spillane and his colleagues' (2009) discussion of organizational routines that ensure regular collective attention to a particular aspect of school work. One such routine, for example, could be a weekly expectation that members of a professional learning community or grade-level team discuss approaches that are in use with all students who teachers believe are not on pace for success. Protocols could help to structure these conversations around high-impact topics. Research suggests, for example, that communication about early warnings is more likely to interrupt the momentum of normal operations and lead to shifts in strategy when it is accompanied by a request for alternate viewpoints (Barton & Sutcliffe, 2009).

Leadership for this aspect of school change, then, involves developing routines that facilitate sharing of teachers' knowledge about emerging learning problems, establishing expectations for the kinds of learning difficulties that should become public within the school, and fostering norms of mutual assistance to make sense of emerging problems.

Orderly transitions

Noticing and discussing problems, as difficult as these might be, are only the first step. What does it take, then, to interrupt the momentum of activity in a classroom or school long enough to consider alternatives? And when should such consideration actually result in a shift in operating mode?

Without well-understood guidelines for when it is appropriate to shift from one approach to another, combining standardized and flexible

strategies could be a recipe for organizational chaos. Can any teacher make this shift at any time, or do certain conditions have to be met first? In the early stages of problem development, it may not be obvious that major changes are needed, but waiting too long for managerial approval can delay needed changes and complicate recovery.

Rules and routines are clearly needed to clarify when a shift to improvisational operations is appropriate, who can reach that decision, and what decision premises should provide guidance. And, because of the frequency with which learning problems are encountered in schools, these routines must be efficient, allowing the school to address multiple threats to reliability simultaneously while sustaining normal operations for other aspects of the work. In schools, then, shifts in operating modes are an ongoing part of operations, unlike the episodic shifts that are more typical of HROs. In effect, schools need to operate in both modes all the time, as different aspects of the work and different students' challenges require improvisational strategies. The resulting challenge is to keep track of the shifting functions that are being addressed in each mode and communicate these well enough to enable coordination of work, collegial support, and supervision.

Gawande's (2009) analysis of the impact of various types of checklists in improving reliability suggests two further requirements for switching from standardized to flexible operating mode. First, the routine should involve collegial discussion, so that decentralized decision making takes full advantage of the expertise of everyone with relevant knowledge, and, second, that it be accountable, in the sense that progress and results are regularly reviewed.

An ongoing case study of one particularly successful elementary school illustrates how such a routine could operate. Teachers are expected to follow prescribed instructional procedures for normal operations, but they also report that they can begin experimenting with alternatives in order to solve an emerging student-learning problem. They do this, however, *after they have discussed the problem with other members of their grade-level team*. The grade-level team, thus, has two different functions. Most of the team's activity focuses on coordinating the regular work of implementing standard programs. But the established relationships, physical proximity, and shared responsibility also create an efficient context for sharing early concerns about student learning and soliciting advice from peers. The authority to move ahead after this discussion with team members supports more rapid adjustments than would be possible if supervisor approval were required. Also, it adds a measure of collective expertise and peer accountability that would be absent if teachers simply made changes that they felt were needed without consultation.

A theory of action for leading educational change

In the preceding discussion, the four distinctive accomplishments of HROs and the corresponding challenges associated with their use in schools combine to frame several strategies for leading educational change. Here is a summary of the larger theory of action for leading school change that these strategies comprise:

- **Balanced and sustainable priorities.** If school leaders establish priorities for what the school should achieve and avoid through inclusive deliberation and open communication about required trade-offs, the resulting goals will be more likely to be

sustained long enough to build the capabilities, structures, and supportive cultures that allow the school system to succeed.

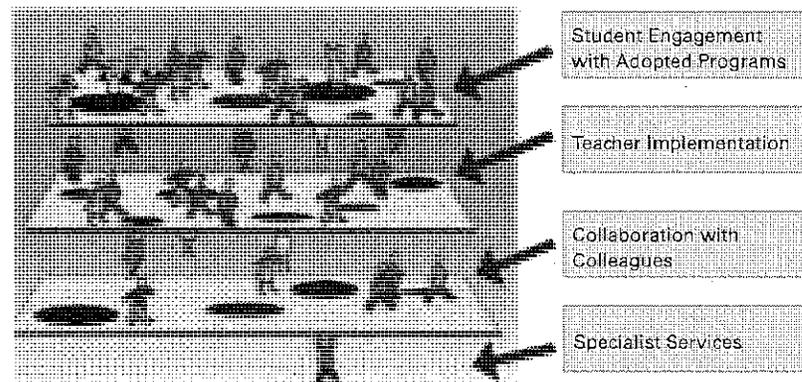
- If schools deliberately develop capacities to work simultaneously in **skeptical standardization** and **constrained improvisation** modes, they will be more likely to achieve both high performance and high reliability.
- **Skeptical standardization.** If standardized procedures are used effectively and skeptically, so that they work with most students, allow slack to deal with exceptions, and are subjected to constant watchfulness for inevitable problems, they will provide an important foundation for reliable performance.
- **Constrained improvisation.** If flexibility for innovation is used with constraints that ensure access to collective expertise and shared accountability for results, it will be a powerful means of addressing the wide variety of problems that emerge as any instructional program is implemented.

- **Public warnings.** If the school creates incentives and routines for sharing information about learning problems in the earliest stages, collective expertise can be brought to bear before those problems cascade into intractable failures.
- **Orderly transitions.** If the school has clear routines and decision frameworks for when operations should shift from skeptical standardization to constrained improvisation, then it is more likely that the strengths of both approaches will be used to achieve goals and avoid failures.

Reason (2000) offers a useful metaphor for how the school's core work of teaching and learning could be shaped by this theory of action for leading school change. He compares HROs to several layers of Swiss cheese, each able to prevent some—but not all—problems from slipping through. As long as holes do not line up, organizations that create several “slices” can effectively prevent failures. Thus, successful HROs construct reliable systems out of several protective layers, each unreliable by itself.

Figure 2

A swiss cheese model of high reliability schooling



Note: Adapted from Bellamy, Fulmer, and Muth (2007). Used with permission of the authors.

Similarly, the challenge of change management in education can be seen as one of constructing and sustaining several layers—program components and strategies—that support a school’s achievement and avoidance goals. Figure 2 illustrates one possibility. Here, the first layer is the adopted curriculum, together with students’ independent engagement with the associated materials. In the present theory of action, this layer is constructed as school leaders develop curriculum frameworks to achieve the school’s priority goals and adopt particular curricula and programs as part of a plan for skeptical standardization. The layer is strengthened as school leaders select research-based programs, ensure access to support materials, and communicate about the importance of the learning objectives. For some students, this is sufficient for learning.

The second layer involves the teacher’s implementation of the standard programs, including explanations, questioning strategies, pacing, and task assignments as well as the more general classroom routines and relationships with students. School leaders support this layer through such activities as professional development and coaching in use of the adopted programs, data systems that provide regular feedback, and opportunities to observe and learn from colleagues who are using the same programs.

The third layer consists of the added resources that a team of colleagues can bring to the task of making sense of student learning problems and responding adaptively when they arise. This layer represents the shift from skeptical standardization to constrained improvisation and offers the possibility that collective teacher expertise can produce effective interventions for some emerging problems. Supporting this layer means fostering diverse

skills and perspectives among teachers, building routines that foster rapid communication about emerging problems, and having clear understandings about when teachers are authorized to shift out of normal procedures.

A final layer involves use of the school’s formal programs for students having difficulty, which typically bring additional resources, expertise, and formal procedures to address continuing difficulties. As a theory of action for leading school change, then, strategies from High Reliability Organizations offer a complex but practical strategy for connecting school organization and management with the core work of teaching and learning.

Conclusion

BP’s tragic accident in the Gulf of Mexico may be relevant for educational leaders for two reasons. First, it highlights just how complex and difficult operating without failures can be, helping to clarify the distinctive capabilities and accomplishments through which HROs achieve accident-free performance. Each of these accomplishments is a major leadership challenge, and each frames a significant agenda for those responsible for educational change.

The second lesson for educators from the BP experience, more implicit in the preceding discussion, lies in the similarities between reports of the company’s operations leading up to the accident and many current conditions in public education. As educators, we confront similar discrepancies between our non-negotiable goals for student learning and the operating cultures of many schools. We often fail to achieve the benefits of standardization with half-hearted implementation of best practices, but then also miss out on the benefits of improvisation due to insufficient support for collaboration.

And we allow early warnings of progress, which are clearly evident to some, to slip by without public discussion and collective action, giving emerging problems time to cascade into intractable failures. Addressing these and other threats to high reliability challenges educators to continue to improve strategies for leading school change.

In sum, the accomplishments of HROs offer a general theory of action for leading school change with equal commitment to what schools should achieve and what they should avoid. Reaching these dual goals requires an ability to operate simultaneously in two operating modes, one centrally controlled and standardized, the other distributed and improvisational. Shared information about emerging problems allows these contrasting modes of operation to be combined in orderly ways. Application of these HRO concepts to educational change is still in an early stage of development, and many details still depend on extrapolation from experience in other settings. Nevertheless, promising results from initial school applications and still-unfulfilled expectations for high-reliability learning offer strong encouragement to continue exploring what schools can learn about avoiding failure from High Reliability Organizations.

References

- Alder, P., Goldoftas, B., & Levine, D. (1999). Flexibility versus efficiency? A case study of model changeovers in the Toyota production system. *Organization Science*, 10, 43–68.
- Barber, M., & Mourshed, M. (2009). *How the world’s best performing schools come out on top*. Washington, DC: McKinsey & Company.
- Barton, M., & Sutcliffe, K. (2009). Overcoming dysfunctional momentum: Organizational safety

- as a social achievement. *Human Relations*, 62(9), 1327–1356.
- Bellamy, T., Crawford, L., Huber-Marshall, L., & Coulter, G. (2005). The fail-safe schools challenge: Leadership possibilities from high reliability organizations. *Educational Administration Quarterly*, 4(3), 383–412.
- Bellamy, T., Fulmer, C., & Muth, R. (2007). Five ideas for reframing the principalship. *Educational Leadership and Administration*, 19, 57–74.
- Bidwell, C. (2001). Analyzing schools as organizations. *Sociology of Education*, 2001 Extra Issue, 100–114.
- Bigley, G., & Roberts, K. (2001). The incident command system: High reliability organizing for highly volatile and complex environments. *Academy of Management Journal*, 44(6), 1281–1299.
- Cameron, K., Quinn, R., DeGraff, J., & Thakor, A. (2006). *Competing values leadership: Creating value in organizations*. Northampton, MA: Edward Elgar.
- Chrislip, D. (2002). *The collaborative leadership fieldbook*. San Francisco, CA: Jossey-Bass.
- Donmoyer, R. (1996). The concept of a knowledge base. In F. Murray (Ed.), *The teacher educator's handbook: Building a knowledge base for the preparation of teachers*. San Francisco, CA: Jossey-Bass.
- Dufour, R., Eaker, R., & Dufour, R. D. (Eds.). (2005). *On common ground: The power of professional learning communities*. Bloomington, IN: National Educational Service.
- Gawande, A. (2009). *The checklist manifesto: How to get things right*. New York, NY: Metropolitan Books.
- Goodlad, J. (1984). *A place called school*. New York, NY: McGraw Hill.
- LaPorte, T. (1996). High reliability organizations: Unlikely, demanding, and at risk. *Journal of Contingencies and Crisis Management*, 4(2), 60–71.
- LaPorte, T., & Consolini, P. (1991). Working in practice but not in theory: Theoretical challenges of high reliability organizations. *Journal of Public Administration Research and Theory*, 1, 19–47.
- Levine, T., & Marcus, A. (2007). Closing the achievement gap through teacher collaboration: Facilitating multiple trajectories of teacher learning. *Journal of Advanced Academics*, 19(1), 116–138.
- Lieberman, A., & Miller, L. (2008). *Teachers in professional communities: Improving teaching and learning*. New York, NY: Teachers College Press.
- Marzano, R., & Waters, T. (2009). *District leadership that works: Striking the right balance*. Bloomington, IN: Solution Tree.
- Perrow, C. (1967). A framework for the comparative analysis of organizations. *American Sociological Review*, 32, 194–208.
- Perrow, C. (1984). *Normal accidents: Living with high risk technologies*. New York, NY: Basic Books.
- Reason, J. (2000, March 18). Human error: Models and management. *British Medical Journal*, 320, 768–770.
- Resar, R. (2006). Practical applications of reliability theory: Making noncatastrophic health care processes reliable. *Health Services Research* 41(4), 1677–1689.
- Roberts, K. (1990). Managing high reliability organizations. *California Management Review*, 32, 101–113.
- Roberts, K., Bea, R., & Bartles, D. (2001). Must accidents happen? Lessons from high-reliability organizations. *Academy of Management Review*, 15(3), 1–12.
- Roberts, K., & Libuster, C. (1993). From Bophal to banking: Organizational design can mitigate risk. *Organizational Dynamics*, 21, 15–26.
- Roberts, K. H., K. Yu, & D. van Stralen (2004). Patient safety as an organizational systems issue: Lessons from a variety of industries. In B. Youngberg and M. Hatlie (Eds.), *Patient safety handbook* (pp. 169–186). Sudbury, MA: Jones and Bartlett.
- Robinson, V., Hoepa, M., & Lloyd, C. (2009). *School leadership and student outcomes: Identifying what works and why*. Auckland, NZ: University of Auckland and New Zealand Ministry of Education.
- Rochlin, G. (1993). Defining “high reliability” organizations in practice: A taxonomic prologue. In K. Roberts (Ed.), *New challenges to understanding organizations* (pp. 11–32). New York, NY: Macmillan.
- Roth, E., Multer, J., & Raslear, T. (2006). Shared situational awareness as a contributor to high reliability performance in rail yard operations. *Organization Studies*, 27(7), 967–987.
- Rowan, B. (1990). Commitment and control: Alternative strategies for the organizational design of schools. In C. Cazden (Ed.), *Review of research in education* (vol. 16, pp. 353–389). Washington, DC: American Educational Research Association.
- Spillane, J., Mesler, L., Croegaert, C., & Sherer, J. (2009). *Coupling administrative practice with the technical core and external regulation: The role of organizational routines*. Evanston, IL: Institute for Policy Research.

Stringfield, S., & Datnow, A. (2002). Systemic supports for schools serving students placed at risk. In S. Stringfield & D. Land (Eds.), *Educating at-risk students* (pp. 269–288). Chicago, IL: National Society for the Study of Education.

Stringfield, S., Reynolds, D., & Schaffer, E. (2010). Improving secondary students' academic achievement through a focus on reform reliability. *School Effectiveness and School Improvement*, 19(4), 409–428.

Stringfield, S., & Yakimowski-Stebnick, M. (2005). Promise, progress, problems, and paradoxes of three phases of accountability: A longitudinal case study of the Baltimore City Public Schools. *American Educational Research Journal*, 42(1), 43–75.

Weick, K. E. (1987). Organizational culture as a source of high reliability. *California Management Review*, 29, 112–127.

Weick, K., & Sutcliffe, K. (2007). *Managing the unexpected: Resilient*

performance in an age of uncertainty (2nd ed.). San Francisco, CA: Jossey-Bass.

York-Barr, J., & Duke, K. (2004). What do we know about teacher leadership? Findings from two decades of scholarship. *Review of Educational Research*, 74(3), 255–316.

Tom Bellamy is professor of education at the University of Washington Bothell. He can be contacted by e-mail at tbellamy@uwb.edu.

Chapter Four

Best in the World: High Performance with High Reliability

By James H. Eck, McREL

Despite low rankings,
a cause for hope

At McREL's Best in the World exploratory gathering (described in Chapter One), Martin West previewed findings from a study conducted by his colleague from Harvard, Paul Peterson, who, along with Eric Hanushek and Ludger Woessmann, compared the mathematics achievement of top-performing students in the United States to that of students in other OECD countries. Once again, the news isn't promising. Using a cross-comparison study calibrating PISA with NAEP results,¹ Hanushek, Peterson, and Woessman (2010) found that no less than 30 of the 56 other countries that participated in the PISA math test had a larger percentage of students who scored at the international equivalent of the advanced level.² Twelve other countries had more than twice the percentage of highly accomplished students as the United States (Hanushek, Peterson, & Woessman, 2010).

In addition, shortly after this October gathering, results from the 2009 PISA were released. Out of 34 countries, the United States ranked 14th in reading, 17th in science and 25th in mathematics (OECD, 2010). Those

scores, although higher than those from 2003 and 2006, still lag far behind the highest scoring countries, including South Korea, Finland, Singapore, China, and Canada. "This is an absolute wake-up call for America," said U.S. Education Secretary Arne Duncan. "The results are extraordinarily challenging to us and we have to deal with the brutal truth. We have to get much more serious about investing in education" (Armario, 2010). Secretary Duncan's voice joins many others who have suggested we have become a nation of sound sleepers when it comes to educating our children.

As bleak as the PISA data may look for the United States, Sir Michael Barber, in his opening presentation at the Best in the World gathering, expressed cause for hope and shared a quote gained from his days as a student at Oxford: British historian George Trevelyan, in his three-volume biography of Garibaldi, wrote, "There come rare moments, hard to distinguish but fatal to let slip, when all must be set upon a hazard" (Barber, 2010). Barber optimistically observed the stars are lined up in the United States, with the new common standards and a push for common assessments, Race to the Top (even with its flaws), and perhaps most

importantly, what he sees as a sense of a national effort to address the achievement gap.

Our focus for the exploratory gathering and this monograph has been expressly on "the other achievement gap" between the United States and other systems of education worldwide. Yet, in light of Schleicher's (2010) conclusion, the variability in achievement gaps found among and within state and district educational systems across the United States demands an equivalent amount of our attention. McKinsey & Company (2009) purport: "In fact, the most striking, poorly understood, and ultimately hopeful fact about the educational achievement gaps in the United States involves the huge differences in performance found between school systems, especially between systems serving similar students" (p. 12).

To once again become among the "best in the world," we at McREL, along with Bellamy and Stringfield, Reynolds, and Schaffer, believe we should not only be looking outward to the highest performing educational systems, but also looking outward to High Reliability Organizations. As noted in Chapter Two, Stringfield, Reynolds, and Schaffer establish

¹ The NAEP scores came from 8th graders in 2005, while PISA 2006 was administered one year later to students at the age of 15, the year at which most American students are in 9th grade.

² While just six percent of U.S. students scored at or above the advanced level cut score on the PISA 2006 exam, 28 percent of Taiwanese students did. At least 20 percent of students in Hong Kong, Korea, and Finland were in the advanced category.

a position that the urgency for high reliability evolves from a realization that: (1) failures of the system have catastrophic results, (2) current levels of performance variability are unacceptable, and (3) much higher levels of performance reliability are possible.

Building a foundation for “true” educational HROs

To get clearer on the concepts, structures, and processes evident in HROs, it is important to first identify the core principles and practices of true High Reliability Organizations. Although research and theory-building on accidents, human performance, and high reliability began earlier (Perrow, 1999; Roberts, 1990), we are grounding our theory in the work of Karl Weick and Kathleen Sutcliffe (2001, 2007). The remainder of this chapter synthesizes their research, the ideas from Stringfield, Reynolds, Schaffer, and Bellamy from the previous chapters, and concludes with McREL’s thinking about how principles, characteristics, and strategies from HROs translate to educational systems.

Stringfield, Reynolds, and Schaffer suggest a “best in the world” education involves two components: (1) knowing what works extremely well, and (2) providing it with remarkable reliability (p. 1). They approach reliability through the lens of school effectiveness research to establish “what works” with the capacity to more reliably deliver it. Their long history of school effectiveness research, coupled with their 11-year study of HRS systems in the United Kingdom, give them a strong basis for their claim. Their results, particularly in the county of Neath-Port Talbot, Wales, indicate that High Reliability Organization principles can be productively applied in school and district contexts.

With their HRS project, Stringfield, Reynolds, and Schaffer were

attempting to merge their abstract understanding of HROs with school effectiveness findings to provide practical guidance to educators. However, in Chapter Three, we saw that Bellamy took a different approach by looking at how schools work through a lens of organizational and change theory. By adding our own research and literature base, McREL is building a theory of action for high-reliability education systems. And our view stems from a perspective gained through conducting several meta-analyses and research syntheses of instruction, school-level effects, extended learning, and school and district leadership. We have most recently synthesized this research into the publication *Changing the Odds for Student Success: What Matters Most* (Goodwin, 2010), in which we present the What Matters Most framework, composed of the following components:

- Guarantee challenging, engaging, and intentional instruction.
- Ensure curricular pathways to success.
- Provide whole-child student supports.
- Create school cultures with high expectations for behavior and learning.
- Develop data-driven, high-reliability systems.

If we know what works, why aren’t we doing it? McREL’s explanation for this is two-fold: (1) what we know about best practice in teaching and in leadership is not being practiced with superior execution, and (2) our educational systems are not well designed to achieve high performance with high reliability. This is both a people problem and a system problem; some suggest it is *mostly* a system problem. High Reliability Organizations recognize that people will make errors and

mistakes, and unless they are intentional, even those should be considered system issues.

Let’s return to our working definition of high-reliability educational systems from Chapter One: *high levels of student performance, achieved as a result of high-quality instruction, delivered through superior execution of effective research-based practices, with low variability in the quality of instruction within and between schools.*

HRO principles and characteristics

Weick and Sutcliffe (2001, 2007) outline five principles of High Reliability Organizations. These five have been interchangeably referred to by the authors and others as principles, hallmarks, and dimensions. We refer to them as principles, and although their names have changed slightly over time, we use them as follows:

1. Preoccupation with failure
2. Reluctance to simplify interpretations
3. Sensitivity to operations
4. Commitment to resilience
5. Organizing around expertise

McREL is using these principles as the foundation for developing a theory of action for high-reliability educational systems. In addition to these principles, there are a few key characteristics of HROs that researchers have identified, as well as a variety of structures, processes, and strategies that specifically transfer to educational systems, as you will see later in this chapter.

Principles of High Reliability Organizations

1. Preoccupation with failure

High Reliability Organizations focus on errors and mistakes. This doesn’t mean they are paralyzed by anxiety about what could go wrong, or that

they fear personal or organizational failure. High Reliability Organizations do, however, adhere to the slogan coined by NASA during the near-catastrophic Apollo 13 mission that “failure is not an option.” There is no acceptable level of loss for a high-reliability organization.

This unwavering attention to the first signs of events that can cascade toward catastrophic failure, or “weak signals” (Weick & Sutcliffe, 2001), positions HROs to respond early and at the source of the problem before it escalates. Just as importantly, HROs do not become complacent with success. Traditional HROs operate continuously under high-risk conditions yet demonstrate safety records approaching 100 percent. Coming close is not acceptable because failure means that lives can be lost.

What if school systems considered student failure as catastrophic as an airplane failing to land safely or a patient failing to recover from surgery? Moreover, what if educators viewed student failure not as the fault of the child, but as a failure of the system? For many, this will require changing core beliefs and assumptions about education. Our standards may have evolved beyond the “sort and select” model of the Industrial Age, but we continue to expect some students to succeed in school and some to fail.

2. Reluctance to simplify interpretations

High Reliability Organizations are highly complex, interconnected systems, technologically and in the amount of human interactions. Humans as a species are very good at finding patterns, but this trait also predisposes us toward categorizing what we observe into what we already know. It subjects us to blind spots where “believing is seeing” (Weick, 2011).

Hoy, Gage, and Tarter (2006) explain that schools need to simplify less and “see” more. Knowing that life in schools is complex, teachers and administrators need to adopt multiple perspectives to understand the shadings that are hidden below the surface of the obvious. While avoiding oversimplification, HROs don’t get so lost in complexity that they do not take action. They utilize sophisticated data collection systems and analysis processes to drill down to the root cause of the problem. They do something and evaluate the response within the system.

Some districts focus their sole attention on a *post-mortem* evaluation of performance on state and national standardized assessments. These data may be helpful for comparing schools and districts and even for program evaluation, but the information comes too late and is of little value for identifying individual student difficulties and responding with real-time intervention.

Other, more reliability-oriented systems use a repertoire of assessments and focus especially on diagnostic and progress monitoring measures in a Response to Intervention strategy. Oliver Wendell Holmes, Jr. nicely sums up this principle with the statement: “I would not give a fig for the simplicity this side of complexity, but I would give my life for the simplicity on the other side of complexity.”

3. Sensitivity to operations

HROs are attentive to the front line, where the real work gets done. (Weick & Sutcliffe, 2007). Sometimes it is referred to as situational awareness, “having the big picture of the moment” (p. 32). This is facilitated by constant interaction and communication throughout the organization, which includes frequent operations meetings, widely distributed real-time measures of

performance, and frequent face-to-face interaction. In schools and school districts, sensitivity to operations may be the guiding principle to drive the effective implementation of professional learning communities (PLCs).

HROs do not allow hierarchies to become dysfunctional bureaucracies (Weick & Sutcliffe, 2001). For Hoy et al. (2006), this principle means staying close to the core function of the organization. For educational systems, the technical core of what we do is teaching and learning. As McKinsey & Company (2007) conclude, it’s all about instruction.

Sensitivity to operations is also about empowering highly competent individuals closest to the event with the ability and responsibility to push the button or throw the switch. Anomalies are noted while they are still tractable and can still be isolated (Weick & Sutcliffe, 2007) and are acted upon before they become a full-blown unexpected event.

4. Commitment to resilience

Despite their best efforts at attending to weak signals of impending failure, HROs do experience failures. However, they also construct multiple preventative measures and containment systems to minimize the effects of accidents, anticipating that the unexpected may happen.

HROs recognize it is impossible to avoid human errors altogether (Weick & Sutcliffe, 2001). They develop capabilities to detect, contain, and bounce back from those inevitable errors that are part of an indeterminate world. When the unexpected happens, the organization rebounds with persistence, resilience, and expertise (Hoy, Gage, & Tarter, 2006). Resilience is that characteristic which encourages people to act while

thinking or acting in order to think more clearly (Weick & Sutcliffe, 2007).

5. Organizing around expertise

HROs cultivate diversity of expertise and perspective. Their focus is on matching expertise with the problem regardless of rank or status (Hoy et al., 2006). Rigid hierarchies have increased vulnerability to errors (Weick & Sutcliffe, 2007). Instead, the decision structure in effective HROs is a hybrid of hierarchy and specialization (Weick & Sutcliffe, 2001). Important decisions are made by important decision makers. The twist, according to Weick and Sutcliffe, is that the designation of who is important migrates to the person or team with acknowledged, problem-specific expertise. This is often a dynamic process, where knowledgeable people self-organize into ad hoc networks to provide expert problem solving. In schools, PLCs should flexibly and adaptively use all the human assets available to them.

Acting with anticipation and containing the unexpected

You probably have gathered from the descriptions of the five principles of HROs that they are highly interconnected. Preoccupation with failure, reluctance to simplify interpretations, and sensitivity to operations together establish a set of principles and repertoire of processes that Weick and Sutcliffe (2001) refer to as “acting with anticipation.” Simply put, HROs work to anticipate the unexpected and prevent small errors and mistakes from occurring in the first place. A commitment to resilience and cultivating deference to expertise enable HROs to contain the unexpected. The HROs Weick and Sutcliffe studied first tried to build in prevention, and then intentionally avoided becoming so complacent that they had prevented all errors.

They tended to adopt organizational mindsets of seeking the early signs of failure and finding remedies quickly.

Beyond the five principles, there are additional characteristics of HROs that deserve attention, particularly because of their potential in developing a theory of action for high reliability in educational systems. One such characteristic is “mindfulness.”

Mindfulness

High Reliability Organizations attend to the five principles through a constant state of *mindfulness*. Weick and Sutcliffe (2001, 2007) expand Langer’s (1989) conception of individual mindfulness to the level of the organization. HROs are characterized by “an underlying style of mental functioning that is distinguished by continuously updating and a deepening of increasingly plausible interpretations of what the context is, what problems define it, and what remedies it contains” (Weick & Sutcliffe, 2001, p. 3). A mindful organization is more than the sum of mindful individuals (Hoy, 2003).

Mindful organizations also manage the unexpected in early stages, when the signals of trouble are subtle and weak. They encourage the reporting of errors and any failure, no matter how small, as a window to the functioning of the system as a whole (Weick & Sutcliffe, 2001), and develop “a rich awareness of discriminatory detail” (p. 32).

Mindful organizations develop and use enabling structures and processes that enable error identification and correction, cooperation, collaboration, innovation, improvisation, and creativity. Conversely, mindless organizations develop and utilize inhibiting structures and processes. Mindlessness is characterized by “a style of mental functioning in

which people follow recipes, impose old categories to classify what they see, act with some rigidity, operate on automatic pilot, and mislabel unfamiliar new contexts as familiar old ones” (Weick & Sutcliffe, 2001, p. 92). Although traditional HROs are hierarchically structured and have tightly coupled processes, they realize the need to be flexible, adaptive, and responsive. Rigid bureaucracies are not conducive to mindfulness; in fact, they may produce a mindless standardization (Hoy, 2003).

A key strategy for encouraging mindfulness is the use of after action reviews (AARs). Senge (2006) calls the Army’s AARs “arguably one of the most successful organizational learning methods yet devised.” Not to be outdone by the Army, the Navy refers to their process as “during action reviews.” Wildland firefighters call their process “lessons-learned reviews.” Stringfield, Reynolds, and Schaffer advise that, in order to maintain an ongoing, multi-level alertness to surprises or lapses, HROs build powerful databases that possess relevance to core goals, rich triangulation on key dimensions, and real-time availability. At McREL, we regularly conduct after action reviews, particularly following large projects or events. The purpose of AARs is to learn as an organization, not to place blame or single out individuals. We ask ourselves three questions: (1) What went right and what went wrong? (2) What did we learn? and (3) How can we use this information to improve?

Key elements of high reliability educational systems

We recognize that educational systems are inherently different from those organizations that have traditionally, and accurately, been identified as demonstrating high reliability. It may be a stretch to think of school districts and schools in

terms of “failure-free” operations, and it may be even more of a stretch to put the HRO principles into practice. Nonetheless, under the umbrella of organizational mindfulness, we believe the following key elements from HROs should be in the formula of consistently high-performing educational systems:

- Focus on a few key goals.
- Establish standard operating procedures (SOPs).
- Design structures and processes for defined autonomy and constrained improvisation.
- Create and maintain safe reporting cultures.

Focus on a few key goals

What should be evident by now is that the overarching philosophy of HROs is a preoccupation with failure, translated into goals that everyone in the system not only can articulate, but practice with unwavering attention. Kathleen Roberts (1990), from her research of flight deck operations on nuclear aircraft carriers, relayed this insight from a lower-ranking deckhand: “This is just a bird farm. The birds come in, they get fed, and they go” (p. 172). Stringfield and colleagues in their HRS research, note the importance of defining a clear and finite set of goals, shared at all organizational levels. They stress the need for these goals to be co-constructed between the researchers/reformers with teachers, school leaders, and school systems.

Similarly, McREL’s research on district-level leadership highlighted the importance of specifying a few *non-negotiable goals*, at the district level, that should include goals for student achievement and instruction. Just as important (and statistically significant) was the need for these goals to be collaboratively developed. Once the non-negotiable goals are established,

a third “responsibility” is alignment with and support of those goals, through all levels of the organization (Marzano & Waters, 2009).

Bellamy refers to these as “balanced and sustainable goals,” which constitute the first of his organizational accomplishments of HROs. In Chapter Three, he explained that HROs hold a dual bottom line—balancing commitments to both safety and productivity. These are translated into goals for achievement, coupled with goals for avoidance. Drawing upon McREL’s findings, Bellamy advised that school goals be honestly open to change as a result of ongoing dialog and political decision making, while being stable enough to provide a foundation for learning goals, performance indicators, and an instructional program.

Establish standard operating procedures

From McREL’s perspective, the highest performing systems in the world establish and accomplish non-negotiable goals for instruction that translate into practice in every classroom. In order to increase the quality of instruction and reduce the variability in instructional quality, they establish clear instructional priorities at the system level, establish a systematic and systemwide approach to instruction, invest in teacher preparation and professional development, and develop strong instructional leadership. In other words, they very carefully develop tighter coupling within the system for curriculum and instruction.

If we know what works from decades of effective teaching and effective schools research, in terms of research-based best instruction, we must ask ourselves, “Then why aren’t we doing those things consistently?” By suggesting standard operating procedures for instruction, we are

not suggesting lock-step adherence to a particular instructional approach. Some districts, it seems, have gone too far with their implementation of curriculum pacing guides to the point where every teacher is expected to be on the same page on the same day. In *Classroom Instruction that Works: Research-based Strategies for Increasing Student Achievement* (Marzano, Pickering, & Pollock, 2001), the authors identified nine categories of instructional strategies that correlate with high levels of student achievement. However, mindlessly employing the strategies will not raise student achievement; teachers must understand and act on how, when, and why to use them.

Stringfield et al. suggest that regularly repeated tasks that are determined to be effective should become Standard Operating Procedures. These SOPs do not only include instructional strategies identified from the effective schools research, but also time-saving efficiency measures and identification/intervention procedures for students who appear at risk of failure. Stringfield et al. are quick to point out, however, that these procedures must be applied in relation to context and must evolve as circumstances change.

Bellamy calls his second organizational accomplishment “skeptical standardization.” Standardization may serve best when applied to instructional materials, technology, professional development, data systems, and other resources. Standardization is particularly effective in realizing economy and efficiency. As he points out in Chapter Three, an area for the use of standard operating procedures in education is establishing structures and routines that support collective decision-making by teacher teams. Professional learning communities can provide a platform for such

structures and processes to exist. Installing SOPs such as protocols for reviewing student work, monitoring progress of individual students, and collectively responding to the first signs of failure, may provide one of the most promising applications of high-reliability processes.

Design structures and processes for defined autonomy and constrained improvisation

In their study of district-level leadership, Waters and Marzano (2006) discovered a “surprising and perplexing finding: one study in their meta-analysis found that building autonomy was positively correlated to student achievement in the district; but that same study reported that *site-based management* exhibited a negligible or even negative effect on student achievement. The authors resolved these seemingly contradictory findings by coining the term *defined autonomy* (Waters & Marzano, 2006; Marzano & Waters, 2009). The essence of defined autonomy is that “the superintendent provides autonomy to principals to lead their schools, but expects alignment on district goals and use of resources for professional development” (2006, p. 16). One of the associated practices for defined autonomy that surfaced from the research was that superintendents and district staff recognize that a key function is “allowing for and promoting innovation at the school-level within the context of district goals” (p. 16). Defined autonomy actually resides at multiple levels in educational systems, particularly in the relationship of districts to schools and in the balance of school-level goals and procedures with teacher freedom and flexibility in the classroom. Bellamy calls his third organizational accomplishment “constrained improvisation,” which he describes as “a collective process that occurs within constraints of organizational values, collaboration, and previously

established decision-making routines (see p. 28).

The seemingly paradoxical characteristics of defined autonomy and constrained improvisation lie at the heart of mindful educational organizations as they strive toward higher reliability. This characteristic of HROs could be considered the yin to the yang of standardization. It captures the dual operating modes of centralized procedural control and the necessary organizational flexibility to shift decision making to those closest to the action. In order to attain this flexibility, HROs deliberately sustain diverse perspectives and expertise.

HROs often refer to the on-the-ground improvisation as “work-arounds.” While in-the-moment sensitivity to operations is critical to high reliability functioning, work-arounds sometimes can create what is referred to as a *drift away* from standard operating procedures and effective performance. Thus, work-arounds continue to be a concern, even in traditional HROs such as air transportation and chemical safety. In fact, at the most recent International Conference for High Reliability Organizing, how to manage work-arounds was one of the most frequently discussed topics. The concern is the ability of HROs to differentiate between being flexible and “freelancing” too loosely in the moment and to learn from work-arounds—possibly even incorporating new and better practice into standardized procedures.

Almost paradoxically, for schools it may be those structures and processes for determining when to shift from normal operations to improvisation that most needs to be clearly delineated in terms of SOPs. Expanding upon the use of protocols among teacher teams, there could be clear procedures for identifying students at the first indications of

failure. Trigger mechanisms for shifting a response from the teacher to a team with a diversity of expertise could be clearly articulated. Indeed, this is exactly what schools did in the HRS project that Stringfield describes in Chapter Two.

Create and maintain safe reporting cultures (“just” cultures)

HROs create conditions in which employees notice and communicate about early-stage problems that threaten reliability (and safety) and establish incentives for reporting, even if the reporting is done by the individual who made the error or mistake. HROs are constantly concerned with establishing and maintaining safe reporting cultures, or what many of them refer to as “just” cultures. An excellent example of guidelines to create such conditions in the health care profession appears in the “Principles of a Fair and Just Culture” from the Dana Farber Institute in its *Patient Safety Rounds Toolkit* (2004). The seven principles outlined in this document are based upon a core value that “in order to have the greatest impact and achieve the highest level of excellence, staff must be able to speak up about problems, errors, conflicts and misunderstandings in an environment where it is the shared goal to identify and discuss problems with curiosity and respect” (p. 1).

It may very well be those in the organization with an out-of-the-box perspective are best suited to identify the weak signals of impending failure that the rest overlook. For Stringfield and his colleagues, it also means honoring the flaw finders. HROs respect the opinions of even those who find fault, despite being an occasional thorn in leadership’s side.

“Public warnings and orderly transitions” is Bellamy’s fourth HRO accomplishment. Unlike typical HROs,

where problems that threaten reliability are infrequent, Bellamy reminds us that schools face emerging teaching and learning problems all the time. Individual work-arounds and near misses occur constantly, but in too many schools and districts, there is no systematic approach, and lessons learned don't make it to the collective level and become SOPs.

Weick and Sutcliffe (2001) briefly describe the critical yet tentative elements of "credibility" and "trust," both intangible but essential assets. People must feel safe to be able to identify errors in the system, even if they are the ones who commit them. Of course, the organization must follow up on any reports and take some type of action, even if it is simply an acknowledgement. Never should the flaw finders be ostracized or punished.

Lessons from medicine

During the past 10 years, the medical profession has been actively applying lessons from HROs to reduce errors and mistakes that, if left unchecked, can lead to unnecessary patient suffering and death. Weick and Sutcliffe (2007) cite a report on medical errors by the Institute of Medicine that concluded: "[Health care] is very different from a manufacturing process, mostly because of huge variability in patients and circumstances, the need to adapt process quickly, the rapidly changing knowledge base, and the importance of highly trained professionals who must use expert judgment in dynamic settings" (p. 39). Their description similarly applies to educational systems, and as a result, the application of HRO principles, characteristics, and strategies to health care may provide us with guidance.

Several authors, many of whom are physicians and surgeons themselves, have written about ways the medical profession is seeking to improve

their process reliability from doctor to doctor and among hospitals. For instance, in *Better: A Surgeon's Notes on Performance*, Atul Gawande (2007) describes how hospitals maintain a continual focus on the prevention of failure by implementing overlapping protocols to decrease the possibility of mistakes. The health care profession has begun to explore the application of high reliability concepts to areas in which failure of the system does not result in immediate death, but instead on the "slow burning" events such as infection rates, heart disease, obesity, and long-term care.

In both education and health care, we face the constant challenges that arise from the complexity of human beings, physically and cognitively, of the human frailties of doctors, nurses, teachers, and administrators, and of the human interactions between provider and recipient. Resar (2006) identifies four themes in health care settings that help to explain at least a portion of the gap in process reliability:

1. Extreme dependence on hard work and personal vigilance.
2. Focus on mediocre benchmark outcomes rather than process.
3. Great tolerance of provider autonomy.
4. Failure to create systems that are specifically designed to reach articulated reliability goals.

Resar concludes that "the resulting variability in the process of delivering care forces the organization in which these autonomous providers work to develop a supporting infrastructure that is at best marginally effective" (p. 1683). Is it any wonder we in education are experiencing many of the same issues? Optimistically, lessons from High Reliability

Organizations may provide both professions with routes to improvement.

Lessons learned from failure

Some of the most powerful lessons learned from HROs come from failure. Even though NASA adopted the slogan "failure is not an option," the Challenger and Columbia space shuttle disasters still occurred, resulting in the deaths of 13 talented astronauts, one of whom was to be the first teacher in space. Inquiries into the causes for these incidents found that technical failures and cascading human errors and mistakes were compounded by organizational culture.

More recently, we have seen a similar pattern in the BP oil spill (see Chapter Three for Bellamy's analysis). The natural disaster of the Japanese earthquake and tsunami on March 11, 2011, and the near-meltdown of the Fukushima nuclear power plant, which resulted from multiple failures in prevention and in the design of backup systems, is providing an extraordinary learning opportunity for the nuclear power industry. We can hope that the good news will be that not only do the individual organizations that were directly affected learn by these disasters, but so will entire industries.

As Bellamy concludes, organizations most often fail when goals for performance or achievement supersede goals for avoidance of critical errors. The attention to the bottom line for shareholders surpasses attention to safety. The organizational culture shifts from one that encourages error reporting to one that demands compliance and punishes whistle blowers and those with different perspectives. A recent educational equivalent comes to mind: The investigation into the Atlanta Public Schools cheating scandal. Observations from an *Atlanta Journal-*

Constitution article (Vogell, 2011) include these:

- Across Atlanta Public Schools, staff worked feverishly in secret to transform testing failures into successes.
- Teachers and principals erased and corrected mistakes on students' answer sheets.
- Area superintendents silenced whistle-blowers and rewarded subordinates who met academic goals by any means possible.
- Superintendent Beverly Hall and her top aides ignored, buried, destroyed, or altered complaints about misconduct, claimed ignorance of wrongdoing, and accused naysayers of failing to believe in poor children's ability to learn.

James Reason (2000), another physician, tells us that High Reliability Organizations are not immune to adverse events, but they are able to convert these occasional setbacks into enhanced resilience of the system. Whether or not public education systems in the United States are resilient enough to bounce back from repeated events like this is in question.

Some encouraging news—the latest McKinsey study

How the World's Most Improved School Systems Keep Getting Better (Mourshed, Chijoke, & Barber, 2010) examines 20 school systems from around the world, all of which reported significant, sustained, and widespread gains in student performance on international assessment measures, but each at a different stage in its improvement trajectory. The authors of this new report map out a journey of improvement along different stages of the performance spectrum—from poor to fair, fair to good, good to great, and great to excellent.

The report finds that at early stages (i.e., poor to fair, fair to good), the systems dictate “tighter central process control, with scripted standard operating procedures, ‘back to basics’ simplification of production processes, the creation of reliable data on system performance, tighter governance, such as regular reporting and performance reviews, and re-establishing a shared sense of purpose that is cascaded through all levels of the system” (p. 52).

As systems move upward toward good to great, they are characterized by more highly skilled educators. They provide only loose guidelines on teaching and learning processes because peer-led creativity and innovation inside schools becomes the core driver for raising performance. Standard operating procedures are relaxed and the system moves from tighter to looser control. However, at all stages, the systems focus attention on a few non-negotiable key goals.

Once again, this report emphasized the importance of leadership at various levels of the system. To initiate the change toward improvement, leadership transition was necessary, but once the trajectory was established, leadership stability became very important.

A follow-up to the 2007 study of the world's highest performing systems, this report verifies that many of the HRO principles, characteristics, and strategies we have synthesized in this monograph are being implemented in the world's highest performing systems and in those on successful improvement trajectories. We believe that, by understanding the characteristics and adhering to the key principles of high reliability, education can achieve higher levels of performance and lower levels of failure. For leaders and practitioners, the questions become, “Will we achieve reliability at the levels that true HROs operate?” and “Should we commit to anything less?”

Challenging our assumptions about schooling

Jared Diamond, in his book *Collapse: How Societies Choose to Fail or Succeed* (2006), writes: “Perhaps a crux of success or failure as a society is to know which core values to hold on to, and which ones to discard and replace with new values, when times change” (p. 433). At McREL, we have adapted that statement to this: Perhaps the crux of success or failure of American education is for leaders to know which practices to hold on to, which ones to discard, and how to significantly improve execution of effective research-based practices, as times and external demands change.

A bold new initiative

As a result of the feedback from our initial exploratory event in October 2010, McREL is launching a national “best in the world” consortium of leaders from high-performing U.S. schools and districts that, together, will work to reverse the downward slide of U.S. schools by raising both the “floor” and the “ceiling” of student performance. The Network for Innovative Education is an initiative to reduce the achievement gap not only between low- and high-achieving students in the United States but also between the highest performing systems in the United States and the “best in the world.”

References

- Armario, C. (2010, December 7). *Wake-up call: U.S. students trail global leaders*. Retrieved from http://www.msnbc.msn.com/id/40544897/ns/us_news-life/t/wake-up-call-us-students-trail-global-leaders
- Barber, M. (2010, October 28). *International lessons in education reform*. Presented at the Best in the World Exploratory Gathering, Denver, Colorado.

- Dana-Farber Cancer Institute. (2004). Principles of a fair and just culture. *Patient Safety Rounds Toolkit*. Retrieved from <http://www.dana-farber.org/uploadedFiles/Library/adult-care/care-quality-and-safety/patientsafety-resources/patient-safety-rounding-toolkit/principles-of-a-fair-and-just-culture.pdf>
- Diamond, J. (2006). *Collapse: How societies choose to fail or succeed*. London: Penguin Books.
- Gawande, A. (2007). *Better: A surgeon's notes on performance*. New York, NY: Picador.
- Goodwin, B. (2010). *Changing the odds for student success: What matters most*. Denver, CO: Mid-continent Research for Education and Learning (McREL).
- Hanushek, E. A., Peterson, P. E., & Woessmann, L. (2010, November). *U.S. math performance in global perspective: How well does each state do at producing high achieving students?* Retrieved from http://www.hks.harvard.edu/pepg/PDF/Papers/PEPG10-19_HanushekPetersonWoessmann.pdf
- Hoy, W. K. (2003). An analysis of enabling and mindful school structures: Some theoretical research and practical considerations. *Educational Administration Quarterly*, 41(1), 87-108.
- Hoy, W. K., Gage, C. Q., & Tarter, C. J. (2006). School mindfulness and faculty trust: Necessary conditions for each other? *Educational Administration Quarterly*, 42(2), 236-255.
- Langer, E. J. (1989) *Mindfulness*. Reading, MA: Addison-Wesley/Addison Wesley Longman.
- Marzano, R. J., Pickering, D. J., & Pollock, J. E. (2001). *Classroom instruction that works: Research-based strategies for increasing student achievement*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Marzano, R., & Waters, T. (2009). *District leadership that works*. Bloomington, IN: Solution Tree.
- McKinsey & Company. (2007). *How the world's best performing school systems come out on top*. New York, NY: Author.
- McKinsey & Company. (2009, April). *The economic impact of the achievement gap in America's schools*. New York, NY: Author.
- Mourshed, M., Chijioke, C., & Barber, M. (2010). *How the world's most improved school systems keep getting better*. New York, NY: McKinsey & Company.
- OECD. (2010). PISA 2009 results: *What students know and can do—Student performance in reading, mathematics and science. (Volume 1)*. Retrieved from http://www.oecd.org/document/53/0,3746,en_3225235_1_46584327_46584821_1_1_1_1,0.html
- Perrow, C. (1999). *Normal accidents: Living with high risk technologies*. Princeton, NJ: Princeton University Press.
- Reason, J. (2000, March 18). Human error: Models and management. *British Medical Journal*, 320, 768-770.
- Resar, R. (2006). Practical applications of reliability theory: Making noncatastrophic health care processes reliable. *Health Services Research*, 41(4), 1677-1689.
- Roberts, K. H. (1990). Some characteristics of one type of high reliability organization. *Organization Science*, 1(2), 160-176.
- Schleicher, A. (2010). *The high cost of low educational performance*. [Video]. Available from <http://eux.tv/2010/01/andreas-schleicher-on-the-high-cost-of-low-educational-performance/>
- Senge, P. (2006). *The fifth discipline: The art and practice of the learning organization* (rev. ed.). New York, NY: Doubleday.
- Vogell, H. (2011, July 6) Investigation into APS cheating finds unethical behavior across every level. *Atlanta Journal-Constitution*. Retrieved from <http://www.ajc.com/news/investigation-into-aps-cheating-1001375.html?printArticle=y>
- Waters, J. T., & Marzano, R. J. (2006). *School district leadership that works: The effect of superintendent leadership on student achievement*. Denver, CO: Mid-continent Research for Education and Learning.
- Weick, K. E. (2011, April 20). *Putting HRO into practice*. Presented at High Reliability Organizing Conference, Washington, DC. Retrieved from http://www.high-reliability.org/Presentations/Conferences/Washington_DC/Weick_Hro_2011_revapril7_outline_and_speech.pdf
- Weick, K. E., & Sutcliffe, K. M. (2001). *Managing the unexpected: Assuring high performance in an age of complexity*. San Francisco, CA: Jossey-Bass.
- Weick, K. E., & Sutcliffe, K. M. (2007). *Managing the unexpected: Resilient performance in an age of uncertainty*. San Francisco, CA: Jossey-Bass.

Jim Eck is a senior director in McREL's Field Services. His expertise is in leadership, standards, assessment, and facilitating change efforts within state and local education agencies. He can be contacted at jeck@mcrel.org or 303.632.5517.

MREL

Mid-Continent Research for Education and Learning

4501 ETC BLVD, STE 500, DENVER, CO 80231-2596

T 303.733.4160 • F 303.733.3000

Website: www.mrel.org • Email: info@mrel.org